

```

/rpt_family="MaLR"
62716 .62764
/rpt_family="ID"
65856 .65793
/rpt_family="Alu"
65929 .66103
/rpt_family="MIR"
67019 .67319
/rpt_family="MER121"
67684 .67898

Query Match      100.0%; Score 25; DB 10; Length 178416;
Best Local Similarity 100.0%; Pred. No. 1,2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAGTCCCAATAAACAACACTGTC 25
    |||||
Db 38485 CAGTCCCAATAAACAACACTGTC 38509

RESULT 4
AC101961
LOCUS      AC101961      162337 bp      DNA      linear      HTG 16-JUL-2003
DEFINITION Mus musculus clone RP24-212A21, WORKING DRAFT SEQUENCE, 12
            unordered pieces.
ACCESSION  AC101961
VERSION    AC101961.3 GI:32813670
KEYWORDS   HTG; HIGS; PHASE1; HIGS_DRAFT; HTGS_FULLTOP.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 162337)
            Birren,B., Nusbaum,C. and Lander,E.
            Mus musculus, clone RP24-212A21
            Unpublished
            2 (bases 1 to 162337)

REFERENCE
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Glende,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lebecky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,M., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 162337)
            Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeRellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kanat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,

```

```

TITLE
JOURNAL
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 117780: contig of 117780 bp in length
* 117781 117880: gap of 100 bp
* 117881 118327: contig of 447 bp in length
* 118328 118427: gap of 100 bp
* 118428 119101: contig of 674 bp in length
* 119102 119201: gap of 100 bp
* 119202 119953: contig of 752 bp in length
* 119954 120053: gap of 100 bp
* 120054 121224: contig of 1171 bp in length
* 121225 121324: gap of 100 bp
* 121325 122334: contig of 910 bp in length
* 122335 122334: gap of 100 bp
* 122335 123132: contig of 798 bp in length
* 123133 123232: gap of 100 bp
* 123233 124225: contig of 993 bp in length
* 124226 124325: gap of 100 bp
* 124326 125357: contig of 1032 bp in length
* 125358 125457: gap of 100 bp
* 125458 126682: contig of 1225 bp in length
* 126683 126782: gap of 100 bp
* 126783 129982: contig of 3200 bp in length
* 129983 130082: gap of 100 bp
* 130083 162337: contig of 32255 bp in length.
*
* Location/Qualifiers
*     1.162337
*     /organism="Mus musculus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10090"
FEATURES
source

```

```

Macedonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ranasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 16, 2003 this sequence version replaced gi:22381391.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17795
Center clone name: 212 A.21
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159608 bases at least Q40
Consensus quality: 160658 bases at least Q30
Consensus quality: 161096 bases at least Q20
Insert size: 151000; agarose-fp
Quality coverage: 161237; sum-of-contigs
Quality coverage: 11.3 in Q20 bases; agarose-fp
Quality coverage: 10.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 117780: contig of 117780 bp in length
* 117781 117880: gap of 100 bp
* 117881 118327: contig of 447 bp in length
* 118328 118427: gap of 100 bp
* 118428 119101: contig of 674 bp in length
* 119102 119201: gap of 100 bp
* 119202 119953: contig of 752 bp in length
* 119954 120053: gap of 100 bp
* 120054 121224: contig of 1171 bp in length
* 121225 121324: gap of 100 bp
* 121325 122334: contig of 910 bp in length
* 122335 122334: gap of 100 bp
* 122335 123132: contig of 798 bp in length
* 123133 123232: gap of 100 bp
* 123233 124225: contig of 993 bp in length
* 124226 124325: gap of 100 bp
* 124326 125357: contig of 1032 bp in length
* 125358 125457: gap of 100 bp
* 125458 126682: contig of 1225 bp in length
* 126683 126782: gap of 100 bp
* 126783 129982: contig of 3200 bp in length
* 129983 130082: gap of 100 bp
* 130083 162337: contig of 32255 bp in length.
*
* Location/Qualifiers
*     1.162337
*     /organism="Mus musculus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10090"
FEATURES
source

```

```

/clone="RP24-212A21"
/clone_lib="RP24-212A21 Male Mouse BAC"
1. .11780
/clone="assembly_fragment"
/clone_end:SP6
vector_side:left"
11781. .118327
/note="assembly_fragment"
118428. .119101
/note="assembly_fragment"
119202. .119553
/note="assembly_fragment"
120054. .121224
/note="assembly_fragment"
121325. .122234
/note="assembly_fragment"
122335. .123132
/note="assembly_fragment"
123233. .124225
/note="assembly_fragment"
124326. .125357
/note="assembly_fragment"
125458. .126682
/note="assembly_fragment"
126783. .129982
/note="assembly_fragment"
130083. .162337
/note="assembly_fragment"

```

ORIGIN

```

Query Match      85.6%; Score 21.4; DB 2; Length 162337;
Best Local Similarity 95.7%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 CAGTCACAAATAAACAACACTGT 23

Db 135188 CAGTCACAAATAAACAACACTGT 135210

RESULT 5

```

AL663052      Mouse DNA sequence from clone RP23-202N14 on chromosome X, complete
LOCUS          sequence.
DEFINITION
ACCESSION     AL663052
VERSION       AL663052.8 GI:19572097
KEYWORDS      HTG.
SOURCE        Mus musculus (house mouse)
ORGANISM

```

REFERENCE

```

1
AUTHORS       Heath, P.
TITLE         Direct Submission
JOURNAL       Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
COMMENT       Cambridge, CB10 1SA, UK. E-mail enquiries:
              hamquery@sanger.ac.uk
              On Mar 21, 2002 this sequence version replaced gi:19168580.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence was finished as follows unless otherwise noted: all
              regions were either double-stranded or sequenced with an alternate
              chemistry or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by at least
              one plasmid subclone or more than one M13 subclone; and the
              assembly was confirmed by restriction digest. The following
              abbreviations are used to associate primary accession numbers given
              in the feature table with their source databases: Emi, ENBL; Sw;

```

COMMENT

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 9728: contig of 9728 bp in length
* 9829: gap of unknown length
* 10896: contig of 1068 bp in length
* 10897: gap of unknown length
* 10997: contig of 64068 bp in length
* 75065: gap of unknown length
* 75164: gap of unknown length
* 120261: contig of 45096 bp in length
* 120360: gap of unknown length
* 120361: contig of 9691 bp in length
* 130052: gap of unknown length
* 130151: contig of 22452 bp in length
* 130152: gap of unknown length
* 152604: contig of 14126 bp in length.
* 152704: 166829: contig of 14126 bp in length.

```

The sequence of the clone was established as a mapping and

SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202N14 is from the RPCI-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6.

FEATURES

```

source
1. .202972
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-202N14"
/clone_lib="RPCI-23"

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 10; Length 202972;
Best Local Similarity 91.7%; Pred. No. 79;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 2 AGTCACAAATAAACAACACTGTCC 25

Db 9211 AGTCCTCAATAAACAACAGTGTC 9234

RESULT 6

```

AC146500/c     166829 bp DNA linear HTG 21-AUG-2003
LOCUS          ***
DEFINITION     Danio rerio clone CH211-130P22, *** SEQUENCING IN PROGRESS ***
ACCESSION     AC146500
VERSION       AC146500.1 GI:34013546
KEYWORDS      HTG; HTGS PHASE2; HTGS ACTIVEFIN.
SOURCE        Danio rerio (zebrafish)
ORGANISM

```

REFERENCE

```

AUTHORS        Myers, R.M.
TITLE          Genomic studies of vertebrate diversity
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 166829)
AUTHORS        Talbot, W.S., Rauch, G.J., Grimwood, J., Dickson, M., Schmutz, J. and
              Myers, R.M.
TITLE          Direct Submission
JOURNAL        Submitted (21-AUG-2003) Stanford Human Genome Center, 975
              California Avenue, Palo Alto, CA 94304, USA

```

COMMENT

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 9728: contig of 9728 bp in length
* 9829: gap of unknown length
* 10896: contig of 1068 bp in length
* 10897: gap of unknown length
* 10997: contig of 64068 bp in length
* 75065: gap of unknown length
* 75164: gap of unknown length
* 120261: contig of 45096 bp in length
* 120360: gap of unknown length
* 120361: contig of 9691 bp in length
* 130052: gap of unknown length
* 130151: contig of 22452 bp in length
* 130152: gap of unknown length
* 152604: contig of 14126 bp in length.
* 152704: 166829: contig of 14126 bp in length.

```

The sequence of the clone was established as a mapping and

sequencing collaboration at the Stanford Genome Evolution Center, funded by the NIH Centers of Excellence in Genomic Science (CEGS) initiative (<http://cegs.stanford.edu>). The clone was isolated from the SAC library CHORI-211 (<http://bacpac.chori.org>).

-----Genome Center

Center: Stanford Human Genome Center
Center Code: shgc
Web site: <http://www-shgc.stanford.edu>

-----Project Information

Center Project Name: 2367
Center clone name: CH211-130P22

-----Summary Statistics

Consensus quality: 165931 bases at least Q40
Consensus quality: 166056 bases at least Q30
Consensus quality: 166078 bases at least Q20
Estimated insert size: 164000; agarose-fp estimation
Estimated insert size: 166122; sum-of-contigs

estimation

Quality coverage: 1.01 in Q20 bases; agarose-fp

estimation

Quality coverage: 1 in Q20 bases; sum-of-contigs

estimation.

Location/Qualifiers
1. 166829
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-130P22"

FEATURES

source

ORIGIN

Query Match 81.6%; Score 20.4; DB 2; Length 166829;
Best Local Similarity 95.5%; Pred.No.1.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GTCCACAAATAAACAACTGTC 24

Db 96903 GTCCACAAATAACACTGTC 96892

RESULT 7

AL954819

LOCUS AL954819 178491 bp DNA linear HTG 30-OCT-2003
DEFINITION Danio rerio clone CH211-1F7, WORKING DRAFT SEQUENCE, 2 unordered pieces.

ACCESSION

AL954819.17 GI:37606232

VERSION

HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

Danio rerio (zebrafish)

SOURCE

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 178491)
Pelam,S.

REFERENCE

1 Direct Submission

AUTHORS

JOURNAL Submitted (28-OCT-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA UK. E-mail enquiries: zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 9, 2003 this sequence version replaced gi:37496382.

COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfsh-help@sanger.ac.uk

----- Project Information

Center project name: zcif7

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 178029 bases at least Q40

Consensus quality: 178183 bases at least Q30
Consensus quality: 178289 bases at least Q20
Insert size: 178391; sum-of-contigs
Insert size: 128908; 35.7% error; agarose-fp
Quality coverage: 7.89x in Q20 bases; sum-of-contigs Quality
coverage: 11.26x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 163111: contig of 163111 bp in length
* 163112 163211: gap of 100 bp
* 163212 178491: contig of 15280 bp in length.

FEATURES

source

Location/Qualifiers
1. 178491
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-1F7"
/clone_lib="CHORI-211"
1. 163111
/note="assembly fragment:03239
fragment_chain:1"
163212. 178491
/note="assembly fragment:03723
fragment_chain:1
clone_end:SP6
vector_side:right"

misc_feature

1. 163111
/note="assembly fragment:03239
fragment_chain:1"
163212. 178491
/note="assembly fragment:03723
fragment_chain:1
clone_end:SP6
vector_side:right"

misc_feature

1. 163111
/note="assembly fragment:03239
fragment_chain:1"
163212. 178491
/note="assembly fragment:03723
fragment_chain:1
clone_end:SP6
vector_side:right"

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 178491;
Best Local Similarity 88.0%; Pred.No.1.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAACTGTC 25

Db 58234 CAGTCCACAAATAAACAACTATCC 58258

RESULT 8

AC127100/c

LOCUS AC127100 222745 bp DNA linear HTG 11-OCT-2002
DEFINITION Rattus norvegicus clone CH230-164M9, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.

ACCESSION

AC127100.2 GI:23664869

VERSION

HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 222745)
Muzny D.Marie, Metzker M.Lee, Abramzon S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Anyalebechi V., Ayagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biswas K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Cesar H., Center A.,
Chacko J., Chavez D., Chen R., Chen R., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,

Genregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, I., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huale, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwa, L., Louis, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahandartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Mallosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackaleme, O., Okwono, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Frankko, C., Plummer, F., Poindestre, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Steed, J., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, C., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 222745)
Worley, K. C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 222745)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 10, 2002 this sequence version replaced gi:21743892.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNPW
Center clone name: CH230-164M9
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 198734 bases at least Q40
Consensus quality: 200599 bases at least Q30

Consensus quality: 201828 bases at least Q20
Estimated insert size: 205186; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 221310: contig of 221310 bp in length
* 221311 221410: Gap of unknown length
* 221411 222745: contig of 1335 bp in length.

FEATURES
source
1..222745
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-164M9"
1..1503
/note="wgs_end_extension
clone_end:77"
3684..4529
/note="clone_boundary
clone_end:77
site:EcoRI
end_sequence:BH362192"

misc_feature
1..1503
/note="wgs_end_extension
clone_end:77"
3684..4529
/note="clone_boundary
clone_end:77
site:EcoRI
end_sequence:BH362192"

misc_feature
1..1503
/note="wgs_end_extension
clone_end:77"
3684..4529
/note="clone_boundary
clone_end:77
site:EcoRI
end_sequence:BH362192"

ORIGIN
Query Match 80.8%; Score 20.2; DB 2; Length 222745;
Best Local Similarity 88.0%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAACACTGTC 25
|||||
Db 68558 CAGCCACAAATAAACAACACTTTC 68534
|||||

RESULT 9
AX655462 1019 bp DNA linear PAT 22-MAR-2003
LOCUS
DEFINITION Sequence 5332 from Patent WO03000898.
ACCESSION AX655462
VERSION AX655462.1 GI:29158276
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.

REFERENCE
1
Chang, H. S., Chen, W., Cooper, B., Glazebrook, J., Goff, S. A., Hou, Y. M., Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G. Plant genes involved in defense against pathogens
Patent: WO 03000898-A 5332 03-JAN-2003;
Syngenta Participations AG (CH)

FEATURES
source
1..1019
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 79.2%; Score 19.8; DB 6; Length 1019;
Best Local Similarity 91.3%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGTCCACAAATAAACAACACTGTC 24
|||||


```

Db      657 AGTCCACAAATTAACAACCTGTC 679

RESULT 10
AF391109 LOCUS
DEFINITION Oryza sativa beta-expansin (EXPB16) gene, partial cds.
ACCESSION AF391109
VERSION AF391109.1 GI:21666631
KEYWORDS
SOURCE
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE
AUTHORS Lee,Y. and Kende,H.
TITLE Expression of expansin genes in rice internodes
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2993)
AUTHORS Lee,Y. and Kende,H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2001) MSU-DOE Plant Research Laboratory, Michigan
State University, East Lansing, MI 48824, USA
FEATURES
source
1. .2993
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
/note="this sequence is from the Monsanto rice genome
database, www.rice-research.org; contig OSM126829"
<2388..>2993
/gene="EXPB16"
join(<2388..2553,2663..2775,2916..>2993)
/gene="EXPB16"
/product="beta-expansin"
join(2388..2553,2663..2775,2916..>2993)
/gene="EXPB16"
/codon_start=1
/product="beta-expansin"
/protein_id="AAW73780.1"
/db_xref="GI:21666632"
/translation="MALAAKLPSIIVAFVALACCVLRSSVASYDHRKLSGWSIGGAT
WYGPAAGSGTDGACGYQGDVGPFPNSMIAAGSPSIYESKGCSCYQVKCSGNPSC
SGKPVTVVLTDLCPGGA"

ORIGIN
Query Match 79.2%; Score 19.8; DB 8; Length 2993;
Best Local Similarity 91.3%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGTCCACAAATAACAACCTGTC 24
|||||
Db 2025 AGTCCACAAATTAACAACCTGTC 2047

RESULT 11
AC116314 LOCUS
DEFINITION Trypanosoma cruzi chromosome 3 clone chimeric cosmid Tcc44h21-2
strain CL Brener, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC116314
VERSION AC116314.1 GI:19718715
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE
1 (bases 1 to 26845)
AUTHORS Andersson,B. and Bontempi,E.J.
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 26845)
AUTHORS Andersson,B. and Bontempi,E.J.

```

```

TITLE JOURNAL
COMMENT
Direct Submission
Submitted (26-MAR-2002) Department of Genetics and Pathology,
Uppsala University, Rudbeck Laboratory, Uppsala SE-751 85, Sweden
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
Location/Qualifiers
1. .26845
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/chromosome="3"
/clone="chimeric cosmid Tcc44h21-2"
/complement(1495..2742)
/gene="Tcc44h21-2.1"
/complement(1495..2742)
/gene="Tcc44h21-2.1"
/note="Predicted by Glimmer and TestCode.
Motifs: PF01066 CDP-alcitol phosphatidyltransferase;
FS00379 CDP_ALCOHOL_P_TRANSF;
8 transmembrane domains predicted by SOSUI and TMHMM2"
/codon_start=1
/product="Tcc44h21-2.1"
/protein_id="AAL96364.1"
/db_xref="GI:19718716"
/translation="MDLFLSVPTDRLPVPQLSYLTHVLTTPYDLVLYYPASWT
PNKVTLFGIFMTLLSSLLLLTGMPPTLFLPFPYATIRPASLGDSDKQWPFQPTPLH
PTTLRLPSSIFPSPGNMLLFTCGILNSLYCADNIDGRLARLAKSCIGILDRGLD
CVTSLSTCVALSFGISFNISLAVVAVATVFSHTLTKYEHIFLWGNFVSDVA
MFFCVVHVIPLFPGLASAKVSPALLAYALPEKHAALLPRYFAIMVLYASQVY
VILDIALKNWMIIRITLALLNLMLGVIPYQTLQVKVDPAGYWGPFSYVAIW
FITVSTSTVTVHILYSHCARLPDPTALAGLLFVWLAFINCPAAGMISVYVWHIA
QILCYVDGLQNRHLEGSKQKVV"
/complement(3458..3874)
/gene="Tcc44h21-2.2"
/complement(3458..3874)
/gene="Tcc44h21-2.2"
/note="Predicted by Glimmer and TestCode"
/codon_start=1
/product="Tcc44h21-2.2"
/protein_id="AAL96365.1"
/db_xref="GI:19718717"
/translation="MLRRTIVPLAKVGAAPNGTGQKQPVRRHLLHKKRQRQSLFA
IANDVYDEKRLRQLNAMLEERLPPRLEHNKANGGEALRSVGLSTCRVWSDAVT
CPPIRSQGSFSTKGIHCTTAHMTTPVYRTKIAYK"
/complement(4307..5155)
/gene="Tcc44h21-2.3"
/complement(4307..5155)
/gene="Tcc44h21-2.3"
/note="Predicted by Glimmer (2nd Met) and TestCode.
blastP: 31% identical to gi|19114099|ref|NP_593187.1|
(NC_003424) dna-directed rna polymerase i if and iii 24 kd
polypeptide [Schizosaccharomyces pombe]; 27% to
gi|13812223|ref|NP_113354.1|(NC_002752) RNA polymerase I,
II and III 24.3 kDa subunit [Guillardia theta]; 25% to
gi|15228724|ref|NP_188871.1|(NM_113130) RNA polymerase I,
II and III 24.3 kDa subunit [Arabidopsis thaliana]; etc.
Motifs: PF01191 RNA polymerases H / 23 kDa subunit.
EST hits: gi|4827415|gb|AI668107.1|AI668107 TENG0983 T.
Cruzi epimastigote normalised cDNA Library Trypanosoma
cruzi cDNA clone n582.r 5;
gi|3017817|gb|AA830938.1|AA830938 TENS0788 T. cruzi
epimastigote normalised cDNA Library Trypanosoma cruzi
cDNA clone 788.5'
/codon_start=1
/product="Tcc44h21-2.3"

```


Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
 Submitted (08-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 86992)

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, B., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
 Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 8, 2003 this sequence version replaced gi:27372474.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: l28311
 Center clone name: 2182_D_19

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 47027: contig of 47027 bp in length
 * 47028 47127: gap of 100 bp
 * 47128 86992: contig of 39865 bp in length.

FEATURES
 Location/Qualifiers
 1..86992
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="CTD-2182D19"
 /clone_lib="CITD1 Human BAC"

ORIGIN
 Query Match 79.2%; Score 19.8; DB 2; Length 86992;

Best Local Similarity 91.3%; Pred. No. 2.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAACACTGT 23
 |||||
Db 61466 CAGTCCCAATAAACAACACTGT 61444

RESULT 13
HUMNEUROF 100849 bp DNA linear PRI 20-SEP-1995
LOCUS
DEFINITION
 Human oligodendrocyte myelin glycoprotein (OMG) exons 1-2;
 neurofibromatosis 1 (NF1) exons 28-49; ecotropic viral integration
 site 2B (EVI2B) exons 1-2; ecotropic viral integration site 2A
 (EVI2A) exons 1-2; adenylate kinase (AK3) exons 1-2.

ACCESSION L05367.1
VERSION L05367.1 GI:189152
KEYWORDS
 adenylate kinase; ecotropic viral integration site 2A; ecotropic
 viral integration site 2B; neurofibromatosis type 1;
 oligodendrocyte myelin glycoprotein.

SOURCE Homo sapiens (human)
ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 100849)
AUTHORS Cawthon, R.M., Weiss, R.B., Xu, G., Viskochil, D., Culver, M.,
 Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. and
 White, R.

TITLE A major segment of the neurofibromatosis type 1 gene: cDNA
 sequence, genomic structure, and point mutations
JOURNAL Cell 62 (1), 193-201 (1990)
MEDLINE 90304909
PUBMED 2114220

REFERENCE 2 (bases 1 to 100849)
AUTHORS Cawthon, R.M., O'Connell, P., Buchberg, A.M., Viskochil, D.,
 Weiss, R.B., Culver, M., Stevens, J., Jenkins, N.A., Copeland, N.G. and
 White, R.

TITLE Identification and characterization of transcripts from the
 neurofibromatosis 1 region: the sequence and genomic structure of
 EVI2 and mapping of other transcripts
JOURNAL Genomics 7 (4), 555-565 (1990)
MEDLINE 90353953
PUBMED 2117566

REFERENCE 3 (bases 1 to 100849)
AUTHORS Cawthon, R.M., Andersen, L.B., Buchberg, A.M., Xu, G.F., O'Connell, P.,
 Viskochil, D., Weiss, R.B., Wallace, M.R., Marchuk, D.A.,
 Culver, M., Stevens, J., Jenkins, N.A., Copeland, N.G., Collins, F.S. and
 White, R.

TITLE cDNA sequence and genomic structure of EVI2B, a gene lying within
 an intron of the neurofibromatosis type 1 gene
JOURNAL Genomics 9 (3), 446-460 (1991)
MEDLINE 91236164
PUBMED 1903357

REFERENCE 4 (bases 1 to 100849)
AUTHORS Viskochil, D., Cawthon, R., O'Connell, P., Xu, G.F., Stevens, J.,
 Culver, M., Carey, J. and White, R.

TITLE The gene encoding the oligodendrocyte-myelin glycoprotein is
 embedded within the neurofibromatosis type 1 gene
JOURNAL Mol. Cell. Biol. 11 (2), 906-912 (1991)
MEDLINE 91117257
PUBMED 1899288

REFERENCE 5 (bases 1 to 100849)
AUTHORS Wallace, M.R., Andersen, L., Letcher, R., Odeh, H., Saulino, A.M.,
 Fountain, J., Brereton, A., Nicholson, J., Mitchell, A.,
 Brownstein, B.H. and Collins, F.

TITLE A de novo Alu insertion results in neurofibromatosis type 1
JOURNAL Nature 353 (6347), 864-866 (1991)
MEDLINE 92043738
PUBMED 1719426

REFERENCE 6 (bases 1 to 100849)
AUTHORS Weiss, R.B., Dunn, D., DiSera, L., Wheatley, W., Kimball, A., Rote, C.,
 Cherry, J., Duval, B., Lee, R., Ferguson, M.W.J. and Gesteland, R.F.

TITLE The Human Neurofibromatosis Type 1 locus: genomic sequence of the

RESULT 14

AC134669 113200 bp DNA linear PRI 02-FEB-2003
 LOCUS Homo sapiens chromosome 17, clone CTD-2370N5, complete sequence.
 DEFINITION AC134669
 ACCESSION AC134669
 VERSION HTG.
 KEYWORDS AC134669.4 GI:28195535
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 113200)
 Birren,B., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens chromosome 17, clone CTD-2370N5

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 113200)

AUTHORS

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (28-SEP-2002) Whitehead Institute/MIT Center for Genome

REFERENCE

Research, 320 Charles Street, Cambridge, MA 02141, USA

AUTHORS

3 (bases 1 to 113200)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
 Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (15-JAN-2003) Whitehead Institute/MIT Center for Genome

REFERENCE

Research, 320 Charles Street, Cambridge, MA 02141, USA

AUTHORS

4 (bases 1 to 113200)
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cooke,P., Cook,P., Corum,B., Dearellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,

Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (02-FEB-2003) Whitehead Institute/MIT Center for Genome

REFERENCE

Research, 320 Charles Street, Cambridge, MA 02141, USA

AUTHORS

On Feb 2, 2003 this sequence version replaced gi:27753749.

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WISR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@genome.wi.mit.edu

Project Information

Center project name: L28272

Center clone name: 2370_N5

Location/Qualifiers

1. 113200

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/map="17"

/clones="CTD-2370N5"

/clone_lib="CITD1 Human BAC"

22. 327

/rpt_family="AluSq"

1062. 1092

/rpt_family="AT rich"

complement(1195. 1275)

/rpt_family="MIR"

1962. 2079

/rpt_family="FLAM_C"

2081. 2221

/rpt_family="(TAGA)n"

2223. 2508

/rpt_family="AluCo"

2516. 2572

/rpt_family="(TAAAAA)n"

2696. 2783

/rpt_family="MIR"

2868. 2890

/rpt_family="AT rich"

complement(2937. 3009)

/rpt_family="HALib"

3177. 3486

/rpt_family="AluSq"

3501. 3755

/rpt_family="L1MSd"

complement(4077. 4166)

/rpt_family="L1ME"

5144. 5216

/rpt_family="MIR"

complement(5996. 6291)

/rpt_family="AluSq"

6761. 6973

/rpt_family="L2"

7407. 7702

/rpt_family="AluSc"

complement(7759. 8065)

/rpt_family="AluJo"

8422. 8536

/rpt_family="L1MC/D"

complement(8551. 9116)

/rpt_family="MLT1K"

```

repeat_region      9387..9701
/rpt_family="AluJb"
complement(9785..9880)
/rpt_family="MIR"
repeat_region      complement(9911..10069)
/rpt_family="MER5A"
9930..10070
/rpt_family="MER5A"
complement(10071..10134)
/rpt_family="MIR"
10310..10530
/rpt_family="AT-rich"
10533..10837
/rpt_family="AluSg"
11836..11856
/rpt_family="AT-rich"
12376..12401
/rpt_family="(CAA)n"
complement(12407..12550)
/rpt_family="MIR3"
14242..14347
/rpt_family="Tigger7"
14349..14715
/rpt_family="Tigger7"
14936..15284
/rpt_family="AluY"
15510..15758
/rpt_family="MER46C"
15785..15892
/rpt_family="MLT1J2"
complement(15893..16321)
/rpt_family="Tigger2a"
16322..16548
/rpt_family="MLT1J2"
16555..16820
/rpt_family="LTR33"
16949..16997
/rpt_family="LTR33"
17026..17181
/rpt_family="FRAM"
17427..17479
/rpt_family="T-rich"
17544..17571
/rpt_family="(T)n"
complement(17645..17954)
/rpt_family="AluSg"
18006..18297
/rpt_family="AluJc"
18351..18392
/rpt_family="5S"
18540..18653
/rpt_family="FLAM A"
complement(19172..19286)
/rpt_family="MER58C"
19287..19580
/rpt_family="AluSg"
complement(19581..19676)

Query Match      79.2%; Score 19.8; DB 9; Length 113200;
Best Local Similarity 91.3%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGTCCACAATAAACAACACTGT 23
|||||
Db 60048 CAGTCCACAATAAACAACACTGT 60070

RESULT 15
AP005289
LOCUS      139503 bp DNA linear HTG 31-MAY-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone
OJ1112_F09, *** SEQUENCING IN PROGRESS ***
ACCESSION AP005289

VERSION AP005289.1 GI:21280339
KEYWORDS HTG; HTGS PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone:OJ1112_F09
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 139503)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/)
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..139503
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="OJ1112_F09"

ORIGIN
Query Match      79.2%; Score 19.8; DB 2; Length 139503;
Best Local Similarity 91.3%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGTCACACAATAAACAACACTGTC 24
|||||
Db 90916 AGTCACACAATAAACAACACTGTC 90938

Search completed: March 25, 2004, 12:49:54
Job time : 175.338 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 09:53:09 ; Search time 188.926 Seconds

(without alignments)
3951.570 Million cell updates/sec

Title: US-09-963-285-1_COPY_378_402

Perfect score: 25

Sequence: 1 cagtcacacataaacaactgtcc 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_estc.*

10: gb_est2.*

11: gb_estc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_nam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rtd.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Result

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	644	28	B2249455
C 2	19.4	77.6	471	10	BB690229
C 3	19.4	77.6	750	13	CA085229
C 4	19.4	77.6	804	13	EX075142

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	19.4	77.6	1109	28	CC277744
6	19.2	76.8	490	28	AZ027906
7	19.2	76.8	502	28	AQ838510
8	19.2	76.8	640	29	CE568205
9	19.2	76.8	707	28	BZ413252
10	19.2	76.8	857	28	AZ548099
11	19.2	76.8	940	28	AZ536025
12	18.8	75.2	417	14	T75397
13	18.8	75.2	476	9	AI501346
14	18.8	75.2	476	28	B47134
15	18.8	75.2	664	12	BI395174
16	18.8	75.2	930	10	BE973161
17	18.6	74.4	446	14	CF214662
18	18.6	74.4	456	14	CF214741
19	18.6	74.4	479	14	CB346046
20	18.6	74.4	494	12	BI803368
21	18.6	74.4	552	28	AQ652543
22	18.6	74.4	570	13	BQ985897
23	18.6	74.4	594	14	CF114153
24	18.6	74.4	602	13	BQ795615
25	18.6	74.4	609	14	CB916135
26	18.6	74.4	623	14	CD013354
27	18.6	74.4	626	28	AQ953684
28	18.6	74.4	636	28	B94941
29	18.6	74.4	641	13	BQ796004
30	18.6	74.4	663	14	CB981751
31	18.6	74.4	721	29	CE282804
32	18.6	74.4	724	14	CF405984
33	18.6	74.4	764	28	BH052648
34	18.6	74.4	774	14	CD012425
35	18.6	74.4	785	13	BQ797479
36	18.6	74.4	793	14	CF763774
37	18.6	74.4	821	13	BQ797186
38	18.6	74.4	826	14	CB981682
39	18.6	74.4	881	13	BQ797594
40	18.6	74.4	902	13	BQ269957
41	18.6	74.4	945	29	CG047759
42	18.6	74.4	1173	28	CC195469
43	18.4	73.6	173	14	CB931753
44	18.4	73.6	242	9	AA650317
45	18.4	73.6	334	28	BZ652437

ALIGNMENTS

RESULT 1
BZ249455/c
LOCUS
DEFINITION
CH230-314P20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-314P20, genomic survey sequence.
ACCESSION
BZ249455
VERSION
BZ249455.1 GI:23910687
KEYWORDS
GSS.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 644)
Zhuo.S., Shetty,J., Shatsman,S., Tseng,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 Mbol segment
Unpublished (1999).
Other GSSs: CH230-314P20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org

BZ249455 644 bp DNA linear GSS 12-OCT-2002
CH230-314P20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-314P20, genomic survey sequence.

ACCESSION
BZ249455
VERSION
BZ249455.1 GI:23910687
KEYWORDS
GSS.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 644)
Zhuo.S., Shetty,J., Shatsman,S., Tseng,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 Mbol segment
Unpublished (1999).
Other GSSs: CH230-314P20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org

Clones are derived from the rat BAC library CHORI-230 (<http://www.chori.org/bacpac/rat230.html>). For BAC library availability, please contact Pieter de Jong (pieterjong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/orering_information.html). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 314 row: P column: 20

Seq primer: SP6

Class: BAC ends

Location/Qualifiers

```
1. .644
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SnHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-314P20"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SnHsd/MCW) BAC library produced by
Pieter de Jong"
```

ORIGIN

Query Match 100.0%; Score 25; DB 28; Length 644;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAACTGTCC 25

DB 162 CAGTCCACAAATAAACAACTGTCC 138

RESULT 2

BB690229

LOCUS

DEFINITION BB690229 RIKEN full-length enriched, 12 days embryo female mullerian duct Mus musculus cDNA clone 6820446011 3', mRNA sequence.

ACCESSION BB690229.1

VERSION BB690229.1 GI:16016962

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 471)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuura, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiragawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Watanishi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suihito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

further details.

e mouse tissues.

FEATURES

source

```
1. 471
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6J"
/db_xref="taxon:10090"
/clone="6820446011"
/sex="female"
/tissue_type="mullerian duct includes surrounding region"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 12 days embryo
female mullerian duct"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGATTCGAGTTAATTAATTAATTCCTCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI"
```

ORIGIN

Query Match 77.6%; Score 19.4; DB 10; Length 471;
Best Local Similarity 95.2%; Pred. No. 7.3e-03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAAACT 21

DB 155 CAGTCCACAAATAAACAAACT 175

RESULT 3

CA085229

LOCUS

DEFINITION CA085229 750 bp mRNA linear EST 23-SEP-2003
SCEZAM2096B03.g AM2 Saccharum officinarum cDNA clone SCEZAM2096B03
5', mRNA sequence.

ACCESSION CA085229

VERSION CA085229.1 GI:34938536

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Saccharum.

1 (bases 1 to 750)

Vettore, A. L., da Silva, F. R., Kemper, E. L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

JOURNAL

COMMENT Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 096 row: B column: 03
Seq primer: 77 Promoter Primer.
Location/Qualifiers
1..750

FEATURES

source
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCEZAM2096B03"
/lab_host="DH10B"
/clone_lib="AM2"
/note="Organ: Apical meristem and tissues surrounding of
immature plants; Vector: pSport1; Site_1: SalI; Site_2:
NotI; An unidirectional cDNA library generated from
[apical meristem and tissues surrounding of immature
plants]. cDNA was prepared from polyA+ mRNA using
SuperScript Plasmid System Kit (Invitrogen). The
double-strand cDNAs were fractionated in a Sepharose
CL-2B 40cm-columns and fragments sizing between 0.8 and
1.5 Kb were directionally cloned into the vector. Details
of each source of RNA and library construction can be
obtained at <http://sucsest.lad.ic.unicamp.br/public>"

ORIGIN

Query Match 77.6%; Score 19.4; DB 13; Length 750;
Best Local Similarity 95.2%; Pred. No. 6.3e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CCACAAATAAACAACACTGTC 25
|||||
DB 441 CCACAAATAAACAACACTGTC 461

RESULT 4
BX075142/c 804 bp mRNA linear EST 25-APR-2003
LOCUS
DEFINITION
(tcd) Oncorhynchus mykiss cDNA clone tcd003a.o.23 5prim, mRNA
sequence.
BX075142 1 GI:27738561
BX075142
SOURCE
Oncorhynchus mykiss (rainbow trout)

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 804)
Govoroun, M., Guiguen, Y. and Le Gac, F.
Construction and primary characterization of normalized cDNA
libraries in rainbow trout, *Oncorhynchus mykiss*
Unpublished (2003)
Contact: Guiguen Y

INRA - SCRIBE
Campus de Beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.

Plate: 0003 row: 0 column: 23
Seq primer: T7.

FEATURES

Location/Qualifiers
1..804

/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcd003a.o.23"
/tissue_type="adipose tissue, blood, brain,
differentiating gonads, interrenal, intestine, kidney,
liver, muscle, ovary, pituitary, testis"
/dev_stage="unknown"
/lab_host="DH10B"
/clone_lib="AGENAE Rainbow trout normalized multi-tissues
library (tcd)"
/note="Vector: pYT3D-pac; Clone distribution : AGENAE
Resource Centre. Francois PIUMI,
Francois.PIUMI@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, France"

ORIGIN

Query Match 77.6%; Score 19.4; DB 13; Length 804;
Best Local Similarity 95.2%; Pred. No. 6.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCACAAATAAACAACACTGTC 24
|||||
DB 382 TCACAAATAAACAACACTGTC 362

RESULT 5
CC277744 1109 bp DNA linear GSS 13-MAY-2003
LOCUS
DEFINITION
CH261-95B8.RM1.1 CH261 Gallus gallus genomic clone CH261-95B8,
genomic survey sequence.
CC277744
CC277744 1 GI:30638245
KEYWORDS
SOURCE
ORGANISM

Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 1109)
Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: RM1 TAGGACTCACTATAGGAGA
Class: BAC ends
High quality sequence start: 57
High quality sequence stop: 552.

FEATURES
source
1..1109
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-95B8"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: <http://www.chori.org/bacpac>"

FEATURES

Query Match 77.6%; Score 19.4; DB 28; Length 1109;
Best Local Similarity 95.2%; Pred. No. 5.6e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	4	TCACACAAATAACAAACTGTC	24
Db	1043	TCACACAAATAACAAACTGTC	1063
RESULT 6			
AZ027906			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	1	CAGTCACAAATAACAAACTGTC	24
Db	236	CAGACACAAATAACAAATTGAC	259
RESULT 7			
AQ838510			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	1	AGTCACAAATAACAAACTGTC	25
Db	375	AGTCACAAATAACAAACTATAC	398
RESULT 8			
CE568205			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	2	AGTCACAAATAACAAACTGTC	25
Db	375	AGTCACAAATAACAAACTATAC	398
RESULT 9			
CE568205/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	2	AGTCACAAATAACAAACTGTC	25
Db	375	AGTCACAAATAACAAACTATAC	398
RESULT 10			
CE568205			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	2	AGTCACAAATAACAAACTGTC	25
Db	375	AGTCACAAATAACAAACTATAC	398
RESULT 11			
CE568205			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	2	AGTCACAAATAACAAACTGTC	25
Db	375	AGTCACAAATAACAAACTATAC	398
RESULT 12			
CE568205			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	2	AGTCACAAATAACAAACTGTC	25
Db	375	AGTCACAAATAACAAACTATAC	398
RESULT 13			
CE568205			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM		</	

The dog genome: survey sequencing and comparative analysis

TITLE
JOURNAL
MEDLINE
PubMed

COMMENT

Science 301 (5641), 1898-1903 (2003)
22875432
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

Location/Qualifiers

1. 640
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strains="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

FEATURES
source

ORIGIN

Query Match 76.8%; Score 19.2; DB 29; Length 640;
Best Local Similarity 87.5%; Pred. No. 7.7e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAACTGTC 24

Db 24 CAGTCCACAAATAAACAACTTC 1

RESULT 9

BZ419252

LOCUS

DEFINITION BZ419252 707 bp DNA linear GSS 10-DEC-2002
if51c06.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
if51c06.5', genomic survey sequence.

ACCESSION BZ419252

VERSION BZ419252.1

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished (2002)
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: if51 row: c column: 06
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 707.

FEATURES

source

1. .707
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="if51c06"
/lab_host="DH5a"
/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
/note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
The vector was digested with XbaI and one nucleotide was

added by fill in in the recessive 3' end. The genomic DNA
was reblitzed, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, .b/g reads in pUC19). The same ligation
was transformed into DH5a."

ORIGIN

Query Match 76.8%; Score 19.2; DB 28; Length 707;
Best Local Similarity 87.5%; Pred. No. 7.5e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAGTCCACAAATAAACAACTGTC 24

Db 231 CTGTGCAAAATAAACAACTGTC 254

RESULT 10

AZ548099

LOCUS

DEFINITION AZ548099 857 bp DNA linear GSS 14-NOV-2000
ENTD572TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, genomic survey sequence.

ACCESSION AZ548099

VERSION AZ548099.1

KEYWORDS GSS.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 93
High quality sequence stop: 620.

FEATURES

source

1. .857
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHO81; Site 1: Bat I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 28; Length 857;
Best Local Similarity 87.5%; Pred. No. 7e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGTCCACAAATAAACAACTGTC 25

MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized kidney cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE ID=1792280 The following repetitive elements were found in this cDNA sequence: 1-24, >AT_rich#low_complexity 33-171, >LIMA6#LINE/11 Seq primer: M13 Forward
 POLYA=No.

FEATURES
 source
 1..476
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C2p-ru-f-09-C-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-C2p"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library consisted of a UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

ORIGIN
 Query Match 75.2%; Score 18.8; DB 9; Length 476;
 Best Local Similarity 90.9%; Pred. No. 1.2e+04;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 AGTCCACAAATAAACAAACTGT 23
 179 AGTCCACAAATAAACAAATTAT 200

RESULT 14
 B47134
 LOCUS
 DEFINITION HS-1068-B1-B06-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 278 Col=11 Row=D, genomic survey sequence.
 accession B47134

VERSION B47134.1 GI:2551968
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 476)
 AUTHORS Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
 TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
 JOURNAL Unpublished (1997)
 COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
 University of Washington
 Seattle, WA 98195, USA
 Tel: (206) 616-8744
 Fax: (206) 685-7301
 Email: kzackrone@u.washington.edu
 Sequence Tagged Connector
 Plate: CT 278 row: D column: 11
 Class: BAC ends
 High quality sequence stop: 476.

FEATURES
 Location/Qualifiers
 1..476
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=CT 278 Col=11 Row=D"
 /sex="M"
 /clone_lib="CIT Human Genomic Sperm Library C"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

ORIGIN
 Query Match 75.2%; Score 18.8; DB 28; Length 476;
 Best Local Similarity 90.9%; Pred. No. 1.2e+04;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 CAGTCCACAAATAAACAAACTG 22
 112 CAGCCCCCAATAAACAAACTG 133

RESULT 15
 B1395174
 LOCUS
 DEFINITION B1395174 664 bp mRNA linear EST 17-SEP-2002
 TUDPECL60 5'RACE cDNA Equus caballus CDNA clone TUDPE160, mRNA sequence.
 accession B1395174
 version B1395174.1 GI:16117688
 keywords EST.
 source Equus caballus (horse)
 organism Equus caballus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 REFERENCE 1 (bases 1 to 664)
 AUTHORS Pascual,I., Dhar,A.K., Fan,Y., Paradis,M.R., Arruga,M.V. and Alcivar-Warren,A.
 TITLE Isolation of expressed sequence tags from a thoroughbred horse (Equus caballus) 5'-RACE cDNA library
 JOURNAL Anim. Genet. 33 (3), 231-232 (2002)
 MEDLINE 22027061
 PUBMED 12030932
 COMMENT Contact: Alcivar-Warren, A.
 Department of Environmental and Population Health
 Tufts University School of Veterinary Medicine
 200 Westboro Road, North Grafton, MA 01536, USA
 Tel: (508) 839-7970
 Fax: (508) 839-7091
 Email: acacia.warren@tufts.edu
 The cDNA was isolated from blood cells (buffy coat) of a Thoroughbred septic foal

PCR Primers
 FORWARD: 5'-(CUA)4GCCACGGTCGACTAGTAC-3'
 BACKWARD: 5'-(CAU)4CTGTTGAGCGGATGAGTGAA-3'
 Insert Length: 664 Std Error: 0.00
 Seq primer: M13 reverse.

FEATURES
 source
 1..664
 Location/Qualifiers
 /organism="Equus caballus"
 /mol_type="mRNA"
 /db_xref="taxon:9796"
 /clone="TUDPE160"
 /tissue_type="Blood cells (buffy coat) from a septic female"
 /dev_stage="Foal"
 /lab_host="E. coli DH5"
 /clone_lib="5'RACE cDNA"
 /note="Vector: pAMP1 (Gibco BRL); Obtained by 5'RACE following the protocol of Gibco BRL. cDNA was synthesized using a gene-specific (IL1- α) primer (GSP1: 5'-CTGTTGAGCGGATGAGT-3') from a genomic sequence. dCTP was used to add a homopolymeric tail in the 3'end and the tailed cDNA was amplified by PCR using an anchor IL-1 α gene-specific primer (GSP2: 5'-CAUCAUCAUCTGTTTGGCGGATGAGTGAA-3') as the reverse and a Universal Anchor Primer (Gibco BRL) as the forward primer. cDNA was cloned through UDG cloning method."

ORIGIN

Query Match 75.2%; Score 18.8; DB 12; Length 664;
 Best Local Similarity 90.9%; Pred. No. 1e+04;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGTCACAAATAAACAACTGT 23
 ||| ||||| ||||| |||||
 Db 221 AGTGACAAATAAACAACTGT 242

Search completed: March 25, 2004, 15:30:13
 Job time : 192.926 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 10:03:04 ; Search time 18.1912 Seconds
(without alignments)
4297.861 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423
Perfect score: 21
Sequence: 1 gggattcttagaggaggag 21

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2455946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	6021	9	US-09-963-285-5
2	21	100.0	6458	9	US-09-963-285-1
3	19.4	92.4	575	14	US-10-029-386-2846
C 3	18	85.7	480	15	US-10-027-632-269335
C 5	18	85.7	493	15	US-10-027-632-274752
6	17.8	84.8	9951	16	US-10-257-166-9
C 7	17.4	82.9	185	12	US-10-424-599-98202
C 8	17.4	82.9	507	15	US-10-027-632-190023
C 9	17.4	82.9	507	15	US-10-027-632-190024
C 10	17.4	82.9	507	15	US-10-027-632-190025
11	17.4	82.9	549	9	US-09-734-017A-45
12	16.8	80.0	603	15	US-10-027-632-186
13	16.8	80.0	603	15	US-10-027-632-187
C 14	16.8	80.0	604	15	US-10-027-632-17929
15	16.8	80.0	613	15	US-10-027-632-54479

16	16.8	80.0	613	15	US-10-027-632-54480
17	16.8	80.0	613	15	US-10-027-632-55335
18	16.8	80.0	613	15	US-10-027-632-56157
19	16.8	80.0	613	15	US-10-027-632-56158
20	16.8	80.0	613	15	US-10-027-632-295612
21	16.8	80.0	613	15	US-10-027-632-295613
C 22	16.8	80.0	669	15	US-10-027-632-120651
23	16.8	80.0	766	15	US-10-027-632-55334
24	16.8	80.0	817	15	US-10-027-632-10458
C 25	16.8	80.0	2817	15	US-10-312-273-6
C 26	16.8	80.0	32177	10	US-09-764-891-8062
C 27	16.8	80.0	32177	15	US-10-158-034-113
C 28	16.8	80.0	123025	15	US-10-289-762-1
29	16.4	78.1	355	12	US-10-085-783A-3550
30	16.4	78.1	355	15	US-10-242-535A-3550
C 31	16.4	78.1	422	10	US-09-918-995-8357
32	16.4	78.1	436	14	US-10-060-036-2776
C 33	16.4	78.1	618	15	US-10-027-632-104902
C 34	16.4	78.1	738	15	US-10-027-632-12804
C 35	16.4	78.1	1358	9	US-09-822-830A-212
C 36	16.4	78.1	2805	15	US-10-312-273-8
C 37	16.4	78.1	78785	10	US-09-978-167-3
C 38	16.4	78.1	78953	15	US-10-085-117-31
C 39	16.4	78.1	202001	9	US-09-734-674-3
C 40	16.4	78.1	202001	14	US-10-274-990-3
C 41	16.2	77.1	332	9	US-09-563-817-653
C 42	16.2	77.1	389	14	US-10-313-669-30
C 43	16.2	77.1	479	10	US-09-918-995-31540
C 44	16.2	77.1	550	12	US-10-085-783A-2496
C 45	16.2	77.1	550	15	US-10-242-535A-2496

ALIGNMENTS

RESULT 1
US-09-963-285-5
Sequence-5: Application US/09963285
Patent No. US2002090707A1

GENERAL INFORMATION:
APPLICANT: Enerbck, Sven
APPLICANT: Krook, Katarina
APPLICANT: Rondahl, Lena
APPLICANT: Wasserman, Wyeth
TITLE OF INVENTION: PROMOTER SEQUENCES
FILE REFERENCE: 13425-042001
CURRENT APPLICATION NUMBER: US/09/963,285
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: SE 0004102-0
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US 60/238,897
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: SE 0003435-5
PRIOR FILING DATE: 2000-09-26
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 6021

TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: exon
LOCATION: (1649)...(438)
US-09-963-285-5

Query Match 100.0%; Score 21; DB 9; Length 6021;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGATTCCTAGAGGAGGAG 21

Db 385 GGGATTCCTAGAGGAGGAG 405

RESULT 2
US-09-963-285-1
; Sequence 1, Application US/09963285
; Patent No. US20020090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerbeck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6458
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2235)....(3737)
US-09-963-285-1

Query Match 100.0%; Score 21; DB 9; Length 6458;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
|||||
DB 403 GGGATTCCTAGAGGAGGAG 423

RESULT 3
US-10-029-386-2846/c
; Sequence 2846, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2846
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: Q15046, EVALUE 9.00e-52
; OTHER INFORMATION: NT HIT: D31890.1, EVALUE 1.00e-103
; OTHER INFORMATION: EST_HUMAN HIT: D58129.1, EVALUE 0.00e+00
US-10-029-386-2846

Query Match 92.4%; Score 19.4; DB 14; Length 575;
Best Local Similarity 95.2%; Pred. No. 5.1;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGATTCCTAGAGGAGGAG 21
|||||
DB 88 GGGATTCCTAGAGGAGGAG 68

RESULT 4
US-10-027-632-269935/c
; Sequence 269935, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 269935
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-269935

Query Match 85.7%; Score 18; DB 15; Length 480;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATTCTAGAGGAGGAGGAG 21
|||||
DB 110 ATTCTAGAGGAGGAGGAG 93

RESULT 5
US-10-027-632-274752
; Sequence 274752, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358


```
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274752
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-274752

Query Match      85.7%; Score 18; DB 15; Length 493;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ATTCCTAGAGGAGGAG 21
Db 371 ATTCCTAGAGGAGGAG 388

RESULT 6
US-10-257-166-9
; Sequence 9, Application US/10257166
; Publication No. US2004023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; FILE REFERENCE: 503.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 9
; LENGTH: 9951
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-9

Query Match      84.8%; Score 17.8; DB 16; Length 9951;
Best Local Similarity 90.5%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
Db 9036 GGGATTCGAGAGGAGGAG 9056

RESULT 7
US-10-424-599-98202/c
; Sequence 98202, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 98202
```

```
; LENGTH: 185
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_59690C.1
US-10-424-599-98202

Query Match      82.9%; Score 17.4; DB 12; Length 185;
Best Local Similarity 94.7%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 19
Db 78 GGGATTCCTAGAGGAGGAG 60

RESULT 8
US-10-027-632-190023/c
; Sequence 190023, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190023
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-190023

Query Match      82.9%; Score 17.4; DB 15; Length 507;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
Db 335 GGGATTCCTAGAGGAGGAG 315

RESULT 9
US-10-027-632-190024/c
; Sequence 190024, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190024
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-190024

Query Match 82.9%; Score 17.4; DB 15; Length 507;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCTCTAGGGAAGGAG 21
|||
Db 335 GGGTTTCTCTAGGGAAGCAR 315
|||

RESULT 10
US-10-027-632-190025/c
; Sequence 190025, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190025
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-190025

Query Match 82.9%; Score 17.4; DB 15; Length 507;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCTCTAGGGAAGGAG 21
|||
Db 335 GGGTTTCTCTAGGGAAGCAR 315
|||

RESULT 11
US-09-734-017A-45
; Sequence 45, Application US/09734017A
; Patent No. US20020142422A1

; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; TITLE OF INVENTION: the
; TITLE OF INVENTION: synthesis of amino acids, vitamins, cofactors, nucleotides and
; TITLE OF INVENTION: nucleosides
; FILE REFERENCE: BASF-NAE-1331-99-US
; CURRENT APPLICATION NUMBER: US/09/734,017A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/171,100
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect
; SEQ ID NO 45
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(547)
; OTHER INFORMATION: 85_ppprot1_083_g04
US-09-734-017A-45

Query Match 82.9%; Score 17.4; DB 9; Length 549;
Best Local Similarity 94.7%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCTCTAGGGAAGG 19
|||
Db 247 GGGATTCTCTAGGGAAGG 265
|||

RESULT 12
US-10-027-632-186
; Sequence 186, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 186
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-186

Query Match 80.0%; Score 16.8; DB 15; Length 603;
 Best Local Similarity 90.0%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGGATTCCTAGAGGGAAGCA 20
 ||||| ||||| ||||| ||||| |||||
 Db 532 GGGATTGCTAGAGGGAAGCA 551

RESULT 13

US-10-027-632-187
 ; Sequence 187, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 187
 ; LENGTH: 603
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-187

Query Match 80.0%; Score 16.8; DB 15; Length 603;
 Best Local Similarity 90.0%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGGATTCCTAGAGGGAAGCA 20
 ||||| ||||| ||||| ||||| |||||
 Db 532 GGGATTGCTAGAGGGAAGCA 551

RESULT 14

US-10-027-632-17929/C
 ; Sequence 17929, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 17929
 ; LENGTH: 604
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-17929

Query Match 80.0%; Score 16.8; DB 15; Length 604;
 Best Local Similarity 90.0%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GGGATTCCTAGAGGGAAGCA 21
 ||||| ||||| ||||| ||||| |||||
 Db 302 GGGATTCCTAGAGGGAAGCA 283

RESULT 15

US-10-027-632-54479
 ; Sequence 54479, Application US/10027632
 ; Publication No. US20030204075A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; CURRENT FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 54479
 ; LENGTH: 613
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-54479

Query Match 80.0%; Score 16.8; DB 15; Length 613;
 Best Local Similarity 90.0%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGGATTCCTAGAGGGAAGCA 20
 ||||| ||||| ||||| ||||| |||||
 Db 537 GGGATTGCTAGAGGGAAGCA 556

Search completed: March 25, 2004, 15:52:55
 Job time : 19.1912 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 09:53:09 ; Search time 158.698 Seconds
(without alignments)
3951.570 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423
Perfect score: 21
Sequence: 1 gggattcttagaggaaggag 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vit:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rtd:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	21	100.0	644	28	BZ249455	BZ249455 CH230-314
C 2	19.4	92.4	426	14	D58129	D58129 HUM349D08B
C 3	19.4	92.4	568	12	BM65327	BM65327 UI-E-CK1-
C 4	19.4	92.4	658	13	BU733958	BU733958 UI-E-CK1-

5	19.4	92.4	722	29	CE219158	CE219158 tigr-gss-
6	18.4	87.6	307	10	BB520390	BB520390
7	18.4	87.6	383	10	AW892967	AW892967 CM3-NN000
8	17.8	84.8	241	14	N89865	N89865 zb34h11.s1
9	17.8	84.8	285	10	BB163362	BB163362 BB163362
10	17.8	84.8	287	10	BB333442	BB333442 BB333442
11	17.8	84.8	447	14	CA721082	CA721082 wkmn.pk0
12	17.8	84.8	522	12	BJ483108	BJ483108
13	17.8	84.8	530	14	CA593927	CA593927 wpaic.pk0
14	17.8	84.8	601	13	BQ238208	BQ238208 TaEO5007B
15	17.8	84.8	631	10	BF350523	BF350523 PMO-HT033
16	17.8	84.8	644	13	BQ875250	BQ875250 QG17J08.Y
17	17.8	84.8	698	13	BQ245382	BQ245382 TaE15025C
18	17.8	84.8	720	13	BQ841551	BQ841551 WHE4212.B
19	17.8	84.8	733	29	AG035378	AG035378 Pan trogl
20	17.8	84.8	741	9	AJ613761	AJ613761 AJ613761
21	17.8	84.8	819	10	BF627209	BF627209 HVSMB000
22	17.8	84.8	1260	28	CC227802	CC227802 CH261-162
23	17.4	82.9	596	29	CE582275	CE582275 tigr-gss-
24	17.4	82.9	658	28	B2108019	B2108019 CH230-239
25	17.4	82.9	1009	13	BX363379	BX363379 BX363379
26	17	81.0	597	14	CD727691	CD727691 4032569.1
27	17	81.0	621	29	CE214329	CE214329 tigr-gss-
28	16.8	80.0	197	14	CA735110	CA735110 wpiis.pk0
29	16.8	80.0	216	29	CG516047	CG516047 OST73785
30	16.8	80.0	229	14	CB068446	CB068446 ls19g06.x
31	16.8	80.0	242	10	BF789464	BF789464 602103330
32	16.8	80.0	245	9	AV026659	AV026659 AV026659
33	16.8	80.0	262	9	AI593772	AI593772 vt72d10.x
34	16.8	80.0	271	10	BF784394	BF784394 602110170
35	16.8	80.0	272	12	BG352243	BG352243 sab97g06.
36	16.8	80.0	285	10	BB067618	BB067618 BB067618
37	16.8	80.0	304	9	AA733226	AA733226 vt72d10.r
38	16.8	80.0	304	10	AW236008	AW236008 xn24d06.x
39	16.8	80.0	398	29	CE225854	CE225854 tigr-gss-
40	16.8	80.0	403	13	BY444754	BY444754 BY444754
41	16.8	80.0	424	13	BY444902	BY444902 BY444902
42	16.8	80.0	428	28	BH142762	BH142762 TDGDS43TH
43	16.8	80.0	430	28	AQ218325	AQ218325 HS 2011.A
44	16.8	80.0	433	12	BI963355	BI963355 id27g06.x
45	16.8	80.0	436	28	AZ588700	AZ588700 IM0397N11

ALIGNMENTS

RESULT 1
BZ249455/c
LOCUS
DEFINITION
BZ249455
BZ249455.1 GI:23910687
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 644)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSSs: CH230-314P20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

BZ249455 644 bp DNA linear GSS 12-OCT-2002
CH230-314P20.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-314P20, genomic survey sequence.

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 314 row: P column: 20
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..644
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SENHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-314P20"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 100.0%; Score 21; DB 28; Length 644;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGATTCTAGAGGAGGAG 21

DB 135 GGGATTCTAGAGGAGGAG 115

RESULT 2

D58129/c

LOCUS

DEFINITION HUM349D08B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens
cDNA clone GEN-349D08 5', mRNA sequence.

ACCESSION D58129

VERSION D58129.1

KEYWORDS

SOURCE

ORGANISM

EST.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Teutomu Fujiwara

Otsuka GEN Research Institute

Otsuka Pharmaceutical Co., Ltd

463-10 Kagasuno Kawauchi-cho, Tokushima, 771-01 Japan

Tel: 0886-65-2888

Fax: 0886-37-1035.

FEATURES

source

Location/Qualifiers

1..426
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GEN-349D08"
/clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

ORIGIN

Query Match 92.4%; Score 19.4; DB 14; Length 426;
Best Local Similarity 95.2%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCTAGAGGAGGAG 21

|||||

DB 68 GGGATTCTAGAGGAGGAG 48

RESULT 3

BM665327/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BM665327.1

GI:18972136

EST.

Homo sapiens (human)

1 (bases 1 to 568)

Bonaldi, M.P., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 1-51, >POLY_A\$imple_repeat (matched complement)

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..568
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afl-b-15-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="rhl08 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CK1"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-CK1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bonaldi, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (GT)₁₈ tail. The sequence tag for this
library is GTCC. This library was created for the program,
Gene Discovery in the Visual System, supported by National
Eye Institute (NEI).
TAG_TISSUE=Foveal and Macular Retina
TAG_LIB=UI-E-CK1
TAG_SEQ=GTCC"

ORIGIN

Query Match 92.4%; Score 19.4; DB 12; Length 568;
Best Local Similarity 95.2%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Query Match          92.4%; Score 19.4; DB 13; Length 658;
Best Local Similarity 95.2%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
    |||||
Db 383 GGGATTCCTAGAGGAG 363

RESULT 4
BU733958 658 bp mRNA linear EST 09-OCT-2002
LOCUS UI-E-CK1-af1-b-15-0-UI.s2 UI-E-CK1 Homo sapiens cDNA clone
DEFINITION UI-E-CK1-af1-b-15-0-UI 3', mRNA sequence.
ACCESSION BU733958
VERSION BU733958.1 GI:23661385
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 658)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and substructure: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-51, >POLY A#simple_repeat (matched complement)
Seq primer: M13 FORWARD
POLYA-res.

FEATURES
    source
    1..658
    Location/Qualifiers
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clones="UI-E-CK1-af1-b-15-0-UI"
        /tissue_type="Retina Foveal and Macular"
        /dev_stage="adult"
        /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
        /clone_lib="UI-E-CK1"
        /note="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a
        modified polylinker; Site 1: EcoR I; Site 2: Not I;
        UI-E-CK1 is a normalized cDNA library containing the
        following tissue(s): Retina Foveal and Macular. The
        library was constructed according to Bonaldo, Lennon and
        Soares, Genome Research, 6:791-806, 1996. First strand
        cDNA synthesis was primed with an oligo-dT primer
        containing a Not I site. Double stranded cDNA was ligated
        to an EcoR I adaptor, digested with Not I, and cloned
        directionally into pTV73-Pac vector. The oligonucleotide
        used to prime the synthesis of first-strand cDNA contains
        a library tag sequence that is located between the Not I
        site and the (dT)18 tail. The sequence tag for this
        library is GTCC. This library was created for the program,
        Gene Discovery in the Visual System, supported by National
        Eye Institute (NET).
        TAG_TISSUE=Foveal and Macular Retina
        TAG_LIB=UI-E-CK1
        TAG_SEQ=GTCC"

ORIGIN
Query Match          92.4%; Score 19.4; DB 29; Length 722;
Best Local Similarity 95.2%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
    |||||
Db 406 GGGATTCCTAGAGGAGGAG 426

RESULT 6
BB520390/c
LOCUS BB520390 RIKEN full-length enriched, 16 days neonate heart Mus
DEFINITION musculus cDNA clone D830041M21 3', mRNA sequence.
ACCESSION BB520390
VERSION BB520390.1 GI:9571848
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 307)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,
Karninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Iehli, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

ORIGIN
Query Match          92.4%; Score 19.4; DB 13; Length 658;
Best Local Similarity 95.2%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
    |||||
Db 383 GGGATTCCTAGAGGAG 363

RESULT 5
CE219158 722 bp DNA linear GSS 25-SEP-2003
LOCUS tigr-gss-dog-17000326871021 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE219158
VERSION CE219158.1 GI:35374827
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 722)
Kirkness, E.F., Baina, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.

FEATURES
    source
    1..722
    Location/Qualifiers
        /organism="Canis familiaris"
        /mol_type="genomic DNA"
        /strain="Standard poodle"
        /db_xref="taxon:9615"
        /clone_lib="Dog Library"
        /note="Site 1: BstXI; Libraries were prepared from
        peripheral blood"

ORIGIN
Query Match          92.4%; Score 19.4; DB 29; Length 722;
Best Local Similarity 95.2%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGAG 21
    |||||
Db 406 GGGATTCCTAGAGGAGGAG 426

RESULT 6
BB520390/c
LOCUS BB520390 RIKEN full-length enriched, 16 days neonate heart Mus
DEFINITION musculus cDNA clone D830041M21 3', mRNA sequence.
ACCESSION BB520390
VERSION BB520390.1 GI:9571848
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 307)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,
Karninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Iehli, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

```

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, N., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Oho, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sugabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toyota, T., Tsunoda, Y., Uchihiki, A., Watanabe, S., Yamamura, T., Yamazaki, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suihito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagasaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

Itoh, M., Kiteunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp/>) for further details.

TITLE JOURNAL COMMENT

FEATURES source

1. .307

Location/Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="D83004IM21"

/tissue_type="heart"

/dev_stage="16 days neonate"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 16 days neonate heart"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGGCGCACTCGAGTTTCTTTTCTTAN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTCTTAATTAATTCCTCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 87.6%; Score 18.4; DB 10; Length 307;
Best Local Similarity 95.0%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GGATTCCTAGAGGAGGAG 21
|||||
Db 262 GGATTCCTAGAGGAGGAG 243
|||||

RESULT 7

AW892967/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AW892967 383 bp mRNA linear EST 24-MAY-2000
CM3-NN0006-110300-113-c04 NN0006 Homo sapiens cDNA, mRNA sequence.
AW892967
AW892967.1 GI:8057172
EST.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 383)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Bruneau, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

Qy

Db

RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

Qy

Db

RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

Qy

Db

RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

Qy

Db

RESULT 8

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

RESULT 10
 BB333442/c
 LOCUS
 BB333442 RIKEN full-length enriched, 10 days neonate medulla
 DEFINITION
 oblongata Mus musculus cDNA clone B830012A04 3', mRNA sequence.
 ACCESSION
 BB333442
 VERSION
 BB333442.1 GI:9042205
 EST.
 KEYWORDS
 EST.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Eutelestomii;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 287)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,
 Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,
 Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,
 Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K.,
 Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,
 Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A.,
 Watanabe, S., Yamamura, T., Yamashita, I., Yano, R., Yasunishi, A.,
 Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and
 Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al.)
 UNPUBLISHED (2000)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9226
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagasaka, S.,
 Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kitsuunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tonaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
 Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.
 Location/Qualifiers
 1..287
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="B830012A04"
 /tissue_type="medulla oblongata"
 /dev_stage="10 days neonate"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, 10 days neonate
 medulla oblongata"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGGATCCAGAGGCCTTTTCTTTTCTTTTCTTTT 3'] cDNA was

prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATCTCGAGTTAATTAATTAATTCCTCCCCCCCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 PLC I."

ORIGIN
 Query Match 84.8%; Score 17.8; DB 10; Length 287;
 Best Local Similarity 90.5%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGGATTCTCTAGAGGGAAGGAG 21
 |||||
 DB 91 GGGGTTGCTAGAGGGAAGGAG 71
 |||||
 RESULT 11
 CA721082
 LOCUS
 CA721082 447 bp mRNA linear EST 26-NOV-2002
 DEFINITION
 wkm2n.pk009.m17 wkm2n Triticum aestivum cDNA clone wkm2n.pk009.m17
 5' end mRNA sequence.
 ACCESSION
 CA721082
 VERSION
 CA721082.1 GI:25442875
 KEYWORDS
 EST.
 SOURCE
 Triticum aestivum (bread wheat)
 ORGANISM
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Triticeae; Triticum.
 1 (bases 1 to 447)
 Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
 Miao, G., Caraher, N. and Hanafey, M.K.
 DuPont Wheat cDNA Sequence
 Unpublished (2002)
 Contact: Scott V. Tingey
 Crop Genetics
 E. I. DuPont de Nemours and Company
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
 Tel: 302-631-2602
 Fax: 302-631-2607
 Email: Scott.V.Tingey@usa.dupont.com
 Seq primer: M13.
 Location/Qualifiers
 1..447
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Spring wheat"
 /db_xref="taxon:4565"
 /clone="wkm2n.pk009.m17"
 /tissue_type="kernel"
 /lab_host="DH10B"
 /clone_lib="wkm2n"
 /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
 XhoI; Wheat (Triticum aestivum L.) kernel malted 175 hours
 at 4 C, normalized"

ORIGIN
 Query Match 84.8%; Score 17.8; DB 14; Length 447;
 Best Local Similarity 90.5%; Pred. No. 1.4e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGGATTCTCTAGAGGGAAGGAG 21
 |||||
 DB 120 GGGATACCTAGCGGAAGGAG 140
 |||||
 RESULT 12
 BJ483108/c
 LOCUS
 BJ483108 522 bp mRNA linear EST 23-MAY-2002

```

DEFINITION BJ483108 K. Sato unpublished cDNA library, strain H602 adult,
             heading stage top three leaves Hordeum vulgare subsp. spontaneum
             cDNA clone bah43b10 3', mRNA sequence.
ACCESSION  BJ483108
VERSION     BJ483108.1 GI:21161562
SOURCE      Hordeum vulgare subsp. spontaneum
ORGANISM    Hordeum vulgare subsp. spontaneum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Poideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 522)
AUTHORS     Sato, K., Saisho, D. and Takeda, K.
TITLE       Barley EST sequencing project in NIG and Okayama Univ
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshinigenes.nig.ac.jp.
FEATURES   source
            1..522
            /organism="Hordeum vulgare subsp. spontaneum"
            /mol_type="mRNA"
            /strain="H602"
            /sub_species="spontaneum"
            /db_xref="taxon:77009"
            /clone="bah43b10"
            /cissue_type="top three leaves"
            /dev_stage="adult, heading stage"
            /clone_lib="K. Sato unpublished cDNA library, strain H602
            adult, heading stage top three leaves"
ORIGIN
Query Match      84.8%; Score 17.8; DB 12; Length 522;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCTTAGAGGGAAGGAG 21
    ||||| ||||| ||||| |||||
Db 389 GGGATACCTAGCGGAGGAG 369

RESULT 13
CA593927
LOCUS      CA593927
DEFINITION wpalc.pk004.a21 wpalc Triticum aestivum cDNA clone wpalc.pk004.a21
            5', end, mRNA sequence.
ACCESSION  CA593927
VERSION    CA593927.1 GI:25143946
SOURCE     Triticum aestivum (bread wheat)
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Poideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 530)
AUTHORS     Tingey, S.V., Moore, G., Griffiths, S., Powell, W., Wolters, P.,
            Dolan, M., Hainey, C., Miao, G., Caraher, N. and Hanafey, M.K.
TITLE       DuPont Wheat cDNA Sequence in collaboration with the John Innes
            Center 1
JOURNAL     Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
            Location/Qualifiers
FEATURES
            1..530
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /cultivar="Glenlea"
            /db_xref="taxon:4565"
            /clone="TaE05007B02F"
            /tissue_type="developing seeds"
            /dev_stage="5 days after anthesis"
            /lab_host="E. coli DH10B"
            /clone_lib="TaE05"
            /notes="Vector: pSPORT-P (Invitrogen Technologies); Site 1:
            NotI; Site 2: MluI; mRNA obtained from wheat seeds of
            cultivar Glenlea 5 days post-anthesis"

```

```

source
1..530
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wpalc.pk004.a21"
/tissue_type="anthers"
/lab_host="DH10B"
/clone_lib="wpalc"
/notes="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum) pre-meiotic anthers JIC"
ORIGIN
Query Match      84.8%; Score 17.8; DB 14; Length 530;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCTTAGAGGGAAGGAG 21
    ||||| ||||| ||||| |||||
Db 342 GGGATACCTAGCGGAGGAG 362

RESULT 14
BQ238208/c
LOCUS      BQ238208
DEFINITION TaE05007B02F TaE05 Triticum aestivum cDNA clone TaE05007B02F, mRNA
            sequence.
ACCESSION  BQ238208
VERSION    BQ238208.1 GI:20434084
KEYWORDS   EST.
SOURCE     Triticum aestivum (bread wheat)
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Poideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 601)
AUTHORS     Cloutier, S.
TITLE       Wheat functional genomics - Glenlea developing seeds cDNA libraries
            Unpublished (2002)
JOURNAL     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@agr.gc.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >2.0 kb
            Plate: 007 row: B column: 02
            Seq primer: M13 Forward.
            Location/Qualifiers
FEATURES   source
            1..601
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /cultivar="Glenlea"
            /db_xref="taxon:4565"
            /clone="TaE05007B02F"
            /tissue_type="developing seeds"
            /dev_stage="5 days after anthesis"
            /lab_host="E. coli DH10B"
            /clone_lib="TaE05"
            /notes="Vector: pSPORT-P (Invitrogen Technologies); Site 1:
            NotI; Site 2: MluI; mRNA obtained from wheat seeds of
            cultivar Glenlea 5 days post-anthesis"
ORIGIN
Query Match      84.8%; Score 17.8; DB 13; Length 601;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCTTAGAGGGAAGGAG 21
    ||||| ||||| ||||| |||||
Db 519 GGGATACCTAGCGGAGGAG 499

```

Search completed: March 25, 2004, 15:30:18
Job time : 163.698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 08:39:03 ; Search time 143.084 Seconds
(without alignments)
6361.316 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423

Perfect score: 21

Sequence: 1 gggatccctagagggaaggag 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_hgt:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vl:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_mu:**

20: em_om:**

21: em_or:**

22: em_ov:**

23: em_pat:**

24: em_ph:**

25: em_pl:**

26: em_ro:**

27: em_sts:**

28: em_un:**

29: em_vi:**

30: em_htg_hum:**

31: em_htg_inv:**

32: em_htg_other:**

33: em_htg_mus:**

34: em_htg_pln:**

35: em_htg_rod:**

36: em_htg_mam:**

37: em_htg_vrt:**

38: em_sy:**

39: em_htgo_hum:**

40: em_htgo_mus:**

41: em_htgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	21	100.0	6021	10	MMMFHEAD1	Y08222 M.musculus	
C	21	100.0	168656	9	AC009108	AC009108 Homo sapi	
3	21	100.0	178416	10	AC137554	AC137554 Mus muscu	
C	4	19.4	92.4	156100	2	AC011934	AC011934 Homo sapi
5	19.4	92.4	192590	2	AC118858	AC118858 Rattus no	
6	19.4	92.4	194832	9	AC025287	AC025287 Homo sapi	
7	18.4	87.6	110999	9	AC130893	AC130893 Homo sapi	
8	18.4	87.6	112153	10	EX000520	EX000520 Mouse DNA	
C	9	18.4	87.6	145626	2	AC020721	AC020721 Homo sapi
10	18.4	87.6	173516	2	AP002356	AP002356 Homo sapi	
C	11	18.4	87.6	183298	2	AC025076	AC025076 Homo sapi
12	18.4	87.6	212884	9	AP001541	AP001541 Homo sapi	
13	18.4	87.6	245820	2	AC106477	AC106477 Rattus no	
14	18	85.7	169032	9	AC016931	AC016931 Homo sapi	
C	15	18	187214	2	AC137676	AC137676 Mus muscu	
16	18	85.7	216498	2	AC123722	AC123722 Mus muscu	
17	17.8	84.8	9951	6	AX348314	AX348314 Sequence	
18	17.8	84.8	119185	10	EX255916	EX255916 Mouse DNA	
19	17.8	84.8	144750	9	AL391499	AL391499 Human DNA	
C	20	17.8	84.8	148610	2	AC026175	AC026175 Homo sapi
C	21	17.8	84.8	160184	9	AC015821	AC015821 Homo sapi
22	17.8	84.8	160557	9	AC117945	AC117945 Homo sapi	
C	23	17.8	84.8	160713	2	AC007174	AC007174 Homo sapi
24	17.8	84.8	161300	2	AC026169	AC026169 Homo sapi	
25	17.8	84.8	161449	9	AC005341	AC005341 Homo sapi	
C	26	17.8	84.8	162030	2	AC018601	AC018601 Homo sapi
27	17.8	84.8	165608	2	AC087842	AC087842 Rattus no	
28	17.8	84.8	169199	9	AC018812	AC018812 Homo sapi	
29	17.8	84.8	170977	9	AP000756	AP000756 Homo sapi	
30	17.8	84.8	174896	2	AC022012	AC022012 Homo sapi	
31	17.8	84.8	186064	2	AC020752	AC020752 Homo sapi	
C	32	17.8	84.8	189097	9	AC022002	AC022002 Homo sapi
C	33	17.8	84.8	193423	2	AC026542	AC026542 Homo sapi
C	34	17.8	84.8	213197	2	AC026198	AC026198 Homo sapi
C	35	17.8	84.8	224292	2	AC026873	AC026873 Homo sapi
C	36	17.8	84.8	230956	9	AC090614	AC090614 Homo sapi
C	37	17.8	84.8	231764	2	AC119470	AC119470 Rattus no
C	38	17.8	84.8	240384	2	AC097041	AC097041 Rattus no
39	17.8	84.8	263668	2	AC126697	AC126697 Rattus no	
C	40	17.8	84.8	269916	2	AC132650	AC132650 Rattus no
41	17.4	82.9	565	11	G82157	G82157 S208P6377FA	
C	42	17.4	82.9	84278	2	AC024907	Continuation (4 of
C	43	17.4	82.9	87386	9	AC133750	AC133750 Homo sapi
C	44	17.4	82.9	168373	9	AL391683	AL391683 Human DNA
45	17.4	82.9	177314	2	AC131919	AC131919 Mus muscu	

ALIGNMENTS

RESULT 1
MMMFHEAD1
LOCUS M.musculus MPH-1 gene.
DEFINITION Y08222
ACCESSION Y08222
VERSION Y08222.1 GI:11869968
KEYWORDS mesenchyme fork head-1 protein; MFH-1 gene.
SOURCE mesenchyme (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Miura,N., Iida,K., Kakinuma,H., Yang,X.L. and Sugiyama,T.
TITLE Isolation of the mouse (MFH-1) and human (FKHL14) mesenchyme fork head-1 genes reveals conservation of their gene and protein

MMMFHEAD1 M.musculus MPH-1 gene. 6021 bp DNA linear ROD 14-MAY-1997

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

structures
Genomics 41 (3), 489-492 (1997)
97312712
9169153
2 (bases 1 to 6021)
Miura, N.
Direct Submission
Submitted (18-SEP-1996) N. Miura, Akita University School of
Medicine, Department of Biochemistry, 1-1-1 Hondo, Akita 010, JAPAN
Location/Qualifiers
1. 6021
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129"
/db_xref="taxon:10090"
1486. 4345
2070. 3554
/gene="MPH-1"
2070. 3554
/gene="MPH-1"
/codon_start=1
/product="mesenchyme fork head-1 protein"
/protein_id="CAA69399.1"
/db_xref="GI:1869969"
/db_xref="GOA:Q61850"
/db_xref="SWISS-PROT:Q61850"
/translation="MQARYSVDPNALGVVPLVSEONYRAAGSYGSMASPMGVYSGH
PEQYGMGRSVAPYHOPAPKDLKVPYSYIALITMAIQNAPEKKITLNGLYQPI
DRFFPYRENQWQNSIRHNLSECEVKVPRDDKKPGKGSYWLDPDSYNNFENGSP
LRRRRKKKDVPEKEERAKHLEPSTTAKGAPTGTVPADGPKREKKVYKSEAS
PALPVTIKVETLPEGALQASPSASTPAGSPDGLPEHAAFPNGLPFGSVETIMT
LRTSPGCDLSAAAGAGLVVPLPALPYAAAPAYTQPCAOGLEAGSAGYQCSMRA
MSLYTGAERPAHVCPVPPALDEALSDHPSPGSLGALNLAAGQEGALGASGHHQHG
HLHPQAPPPAPPPAPPPATQATSWLNHGGLSLHPGHTFATQQTFFNVREMFNS
HRLGLDSSLSGESQVSNASQCLPYRATPSLYRAAPYSVDCTKY"
polyA_signal 4320. 4325

ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 6021;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGATTCCTAGAGGAGGAG 21
Db 385 GGGATTCCTAGAGGAGGAG 405

RESULT 2
AC009108/c
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RP11-46309, complete sequence.
AC009108
AC009108.10 GI:24418066
HTG.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 168656)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2 (bases 1 to 168656)
DOE Joint Genome Institute.
Direct Submission
Unpublished
3 (bases 1 to 168656)
Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
3 (bases 1 to 168656)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2800 Mitchell

Drive, Walnut Creek, CA 94598, USA
On Oct 29, 2002 this sequence version replaced gi:13786306.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.snhgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
Location/Qualifiers
1. 168656
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-46309"

ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 168656;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGATTCCTAGAGGAGGAG 21
Db 78330 GGGATTCCTAGAGGAGGAG 78310

RESULT 3
AC127554
LOCUS
DEFINITION Mus musculus BAC clone RP24-323K23 from chromosome 8, complete
sequence.
AC127554
AC127554.4 GI:33457241
HTG.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 178416)
Cordes, M. and Haglund, K.
The sequence of Mus musculus BAC clone RP24-323K23.
Unpublished (2001)
2 (bases 1 to 178416)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 178416)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 178416)
Wilson, R.K.
Direct Submission
Submitted (30-JUL-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 178416)
Wilson, R.K.
Direct Submission
Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 178416)
Wilson, R.
Direct Submission
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 6, 2003 this sequence version replaced gi:33342444.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu

----- Summary Statistics

 Center project name: M_BH0323K23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC124170.

Location/Qualifiers

1. 178416
 /organism="Mus musculus"
 /mol_type="Genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="8"
 /map="8"

/clones="RP24-323K23"
 /clone_lib="RPCI-24"

1791. 1839
 /rpt_family="Alu"

3054. 3288
 /rpt_family="B4"

3834. 3845
 /rpt_family="B4"

3846. 3972
 /rpt_family="Alu"

3973. 4026
 /rpt_family="B4"

6269. 8820
 /note="CpG island (%GC=69.2, o/e=0.78, #CpGs=269)"

16848. 16980
 /rpt_family="MIR"

18383. 18494
 /rpt_family="Alu"

18974. 19317
 /rpt_family="MaLR"

19821. 20013
 /rpt_family="B2"

20188. 20538
 /rpt_family="MaLR"

20539. 20683
 /rpt_family="Alu"

20684. 20799
 /rpt_family="MaLR"

22901. 22957
 /rpt_family="ID"

22968. 23090
 /rpt_family="Alu"

24384. 24441
 /rpt_family="B4"

FEATURES

source

```

repeat_region 25838..26022
               /rpt_family="MER1_type"
repeat_region 27267..27380
               /rpt_family="Alu"
repeat_region 28700..28893
               /rpt_family="B2"
repeat_region 29950..30049
               /rpt_family="B4"
repeat_region 30803..30854
               /rpt_family="ID"
repeat_region 30859..30963
               /rpt_family="Alu"
repeat_region 30987..31062
               /rpt_family="ID"
repeat_region 31157..31294
               /rpt_family="Alu"
repeat_region 32192..32321
               /rpt_family="B4"
repeat_region 32923..33036
               /rpt_family="Alu"
repeat_region 33141..33245
               /rpt_family="Alu"
repeat_region 33394..33521
               /rpt_family="B4"
repeat_region 33806..33891
               /rpt_family="B4"
repeat_region 34643..34976
               /rpt_family="MaLR"
repeat_region 35547..35686
               /rpt_family="B2"
repeat_region 35701..35839
               /rpt_family="Alu"
repeat_region 35873..35982
               /rpt_family="Alu"
misc_feature 39796..41776
               /note="CpG island (%GC=69.3, o/e=0.84, #CpGs=225)"
repeat_region 47196..47354
               /rpt_family="B2"
repeat_region 47365..47498
               /rpt_family="Alu"
repeat_region 47507..47550
               /rpt_family="B4"
repeat_region 47706..47840
               /rpt_family="Alu"
repeat_region 47715..47832
               /rpt_family="B4"
repeat_region 47952..48067
               /rpt_family="Alu"
misc_feature 51423..52857
               /note="CpG island (%GC=66.5, o/e=0.72, #CpGs=121)"
repeat_region 53367..53484
               /rpt_family="Alu"
repeat_region 54128..54233
               /rpt_family="Alu"
repeat_region 55219..55240
               /rpt_family="ERVK"
repeat_region 55241..55364
               /rpt_family="Alu"
repeat_region 55365..55460
               /rpt_family="ERVK"
repeat_region 56094..56379
               /rpt_family="B4"
repeat_region 59534..59702
               /rpt_family="B2"
repeat_region 60089..60165
               /rpt_family="ID"
repeat_region 60513..60709
               /rpt_family="B2"
repeat_region 60886..61042
               /rpt_family="B4"
repeat_region 61852..61997
               /rpt_family="MER1_type"
repeat_region 62247..62549

```

```

repeat_region      /rpt family="MaLR"
62716 .62764
/rpt family="ID"
65655 .65793
/rpt family="Alu"
65929 .66103
/rpt family="MIR"
67019 .67319
/rpt family="MER121"
67684 .67898
repeat_region

```

```

Query Match      100.0%; Score 21; DB 10; Length 178416;
Best Local Similarity 100.0%; Pred. NO. 1.5; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

```

```

QY 1 GGGATTCTAGAGGAGGAG 21
Db 38512 GGGATTCTAGAGGAGGAG 38532

```

```

RESULT 4
AC011934/c
LOCUS          AC011934          156100 bp      DNA      linear      HTG 12-MAR-2000
DEFINITION     Homo sapiens clone RP11-16C11, WORKING DRAFT SEQUENCE, 15 unordered
                pieces.
ACCESSION      AC011934
VERSION        AC011934.5 GI:7230117
KEYWORDS       HTG; HGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 156100)
AUTHORS        Biren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL         Unpublished

```

```

REFERENCE      2 (bases 1 to 156100)
AUTHORS        Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
                Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
                Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
                Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
                Ferreira,P., FitzHugh,V., Forrest,C., Funke,R., Gage,D., Galagan,J.,
                Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
                Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
                Lehoczy,J., Lie,C., Locke,K., Macdonald,P., Marquis,N.,
                McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
                Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
                Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
                Stange-Themann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
                Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL         Direct Submission

```

```

TITLE          Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL         On Mar 12, 2000 this sequence version replaced gi:6970386.
COMMENT         All repeats were identified using RepeatMasker:
                Smit, A.F.A. & Green, P. (1996-1997)
                http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3515
Center clone name: 16_C_11
----- Summary Statistics
Sequencing vector: M13, M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147739 bases at least Q40
Consensus quality: 152170 bases at least Q30
Consensus quality: 153664 bases at least Q20

```

```

Insert size: 163000; agarose-fp
Insert size: 154700; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1060: contig of 1060 bp in length
1061 1160: gap of 100 bp
1161 2774: contig of 1614 bp in length
2775 2875: gap of 100 bp
2876 5568: contig of 2694 bp in length
5569 5669: gap of 100 bp
5670 8685: contig of 3017 bp in length
8686 8785: gap of 100 bp
8786 16002: contig of 7217 bp in length
16003 16102: gap of 100 bp
16103 23355: contig of 7253 bp in length
23356 23455: gap of 100 bp
23456 30787: contig of 7332 bp in length
30788 30888: gap of 100 bp
30889 41339: contig of 10452 bp in length
41340 41439: gap of 100 bp
41440 52179: contig of 10740 bp in length
52180 52766: contig of 10487 bp in length
52767 62866: gap of 100 bp
62867 62867: contig of 13921 bp in length
62868 76887: gap of 100 bp
76888 93898: contig of 17011 bp in length
93899 93999: gap of 100 bp
94000 107072: contig of 13074 bp in length
107073 107173: gap of 100 bp
107174 131104: contig of 23932 bp in length
131105 131204: gap of 100 bp
131205 156100: contig of 24896 bp in length.

```

FEATURES

```

source
1..156100
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="RP11-16C11"
/clone_lib="RP11-16C11" Human Male BAC"
1..1060
/note="assembly_fragment"
1161..2774
/note="assembly_fragment"
2875..5568
/note="assembly_fragment"
5669..8685
/note="assembly_fragment"
8786..16002
/note="assembly_fragment"
16103..23355
/note="assembly_fragment"
23456..30787
/note="assembly_fragment"
30888..41339
/note="assembly_fragment"
41440..52179
/note="assembly_fragment"
52280..62766
/note="assembly_fragment"
62867..76787
/note="assembly_fragment"
76888..93899
/note="assembly_fragment"

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature 93999..107072
/note="assembly_fragment
clone_end:T7
vector_side:right"
107173..131104
/note="assembly_fragment
clone_end:SP6
vector_side:left"
131205..156100
/note="assembly_fragment"

ORIGIN
Query Match 92.4%; Score 19.4; DB 2; Length 156100;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCTTAGGAGGAG 21
Db 152079 GGGATTCTTAGGAGGAG 152059

RESULT 5
AC118858
LOCUS Rattus norvegicus clone CH230-248J11, WORKING DRAFT SEQUENCE, 2
DEFINITION AC118858 linear HTG 15-NOV-2002
unordered pieces.
AC118858
VERSION AC118858.4 GI:25008736
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 192590)
Munzky,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,S., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakolameh,O., Okunonu,G., Olarnpunaagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pul-L.,
Puafo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,J., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 192590)
Worley,K.C.
Direct Submission
Submitted (21-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 192590)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:22856341.
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/) Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWGQ
Center clone name: CH230-248J11
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 165160 bases at least Q40
Consensus quality: 166820 bases at least Q30
Consensus quality: 168087 bases at least Q20
Estimated insert size: 167652; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2625: contig of 2625 bp in length
* 2626 2725: gap of unknown length
* 2726 192590: contig of 189865 bp in length.
Location/Qualifiers
1..192590
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-248J11"
FEATURES
source

```


Homo sapiens X BAC RP13-926M18 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

AC130893
AC130893.3 GI:22657460
HTG.

ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 110999)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Alusbrooks, S.L., Amarutunge, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Binage, K., Blankenbuck, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulaged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Oguchi, M., Okuwon, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scher, S., Scott, G., Shen, H., Shoohtari, N.,
Slason, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vaequez, L., Vera, V., Villaion, D., Vinson, R.,
Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 110999)
Worley, K.C.

Direct Submission
Submitted (15-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 110999)
Worley, K.C.

Direct Submission
Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

4 (bases 1 to 110999)
Worley, K.C.

Direct Submission
Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT
ON Sep 3 2002 this sequence version replaced gi:22538319.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lower quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

```

FEATURES             source
    source
        Location/Qualifiers
            1..110999
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /chromosome="X"
            STS
               11719..11937
                  /standard_name="D11S3059"
            STS
               14741..14954
                  /standard_name="G44369"
            STS
               15930..16107
                  /standard_name="D11S2442"
        misc_feature
               /note="Sized by PCR and restriction digest - 27 bases
               missing."
        misc_feature
               106291..107062
                  /function="force join"
               /note="Sized by PCR and restriction digest - 200 bases
               missing - size of repeat
               971 bases."
        misc_feature
               108513..110999
                  /function="unresolved tandem repeat"
               /note="overlaps bases 1..2481 of clone AC073583"
               /function="clone overlap"
            STS
               108547..1108887
                  /standard_name="D7S3182"

ORIGIN
Query Match      87.6%; Score 18.4; DB 9; Length 110999;
Best Local Similarity 95.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAAGGA 20
    |||||
Db 6684 GGGATTACTAGAGGAAGGA 6703

RESULT 8

```

```

BX000520
LOCUS               112153 bp      DNA      linear      ROD 26-NOV-2003
DEFINITION          Mouse DNA sequence from clone RP23-261N8 on chromosome X, complete
                    sequence.
ACCESSION            BX000520
VERSION              BX000520.12 GI:38564122
KEYWORDS             HTG.
SOURCE               Mus musculus (house mouse)
ORGANISM             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE            1. (bases 1 to 112153)
AUTHORS              Clark, S.
TITLE                Direct Submission
JOURNAL              Submitted (26-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
                    Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                    humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                    On Nov 26, 2003 this sequence version replaced GI:32567731.
COMMENT              Sequence from the Mouse Genome Sequencing Consortium whole genome
                    shotgun may have been used to confirm this sequence. Sequence data
                    from the whole genome shotgun alone has only been used where it has
                    a phred quality of at least 30.
                    ----- Genome Center
                    Center: Wellcome Trust Sanger Institute
                    Center code: SC
                    Web site: http://www.sanger.ac.uk
                    Contact: humquery@sanger.ac.uk
                    -----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-261N8 is from the RPCI-23 Mouse BAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

```

FEATURES             Location/Qualifiers
    source
        1..112153
           /organism="Mus musculus"
           /mol_type="genomic DNA"
           /db_xref="taxon:10090"
           /chromosome="X"
           /clone="RP23-261N8"
           /clone_lib="RPCI-23"

```

ORIGIN

```

Query Match      87.6%; Score 18.4; DB 10; Length 112153;
Best Local Similarity 95.0%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAAGGA 20
    |||||
Db 33551 GGGATTCCTAGAGGAAGGA 33570

RESULT 9
AC020721/c
LOCUS               145626 bp      DNA      linear      HTG 07-JUL-2000

```

Center project name: HumDraft11

Center clone name: CTD-3243P17
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 162824 bases at least Q40
 Consensus quality: 168195 bases at least Q30
 Consensus quality: 170335 bases at least Q20
 Insert size: 171616; sum-of-contigs
 Quality coverage: 5.04x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1      21771 contig of 21771 bp in length
21872 41497 contig of 19626 bp in length
41598 57282 contig of 15685 bp in length
57383 72819 contig of 15437 bp in length
72920 86720 contig of 13801 bp in length
86821 97172 contig of 10352 bp in length
97273 107979 contig of 10707 bp in length
108080 117581 contig of 9502 bp in length
117682 125644 contig of 7963 bp in length
125745 132686 contig of 6942 bp in length
132787 141286 contig of 8500 bp in length
141387 147754 contig of 6368 bp in length
147855 153419 contig of 5359 bp in length
153520 157078 contig of 5211 bp in length
157179 162389 contig of 2445 bp in length
162490 164934 contig of 3212 bp in length
165035 168246 contig of 2239 bp in length
168347 170645 contig of 2239 bp in length
170746 172143 contig of 1398 bp in length
172244 173516 contig of 1273 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      21771: contig of 21771 bp in length
21772 21871: gap of 100 bp
21872 41497: contig of 19626 bp in length
41498 41597: gap of 100 bp
41598 57282: contig of 15685 bp in length
57283 72819: gap of 100 bp
72820 72919: gap of 15437 bp in length
72920 86720: contig of 13801 bp in length
86721 86821: gap of 100 bp
86821 97172: contig of 10352 bp in length
97173 97272: gap of 100 bp
97273 107979: contig of 10707 bp in length
107980 108079: gap of 100 bp
108080 117581: contig of 9502 bp in length
117582 117681: gap of 100 bp
117682 125644: contig of 7963 bp in length
125645 125744: gap of 100 bp
125745 132686: contig of 6942 bp in length
132687 132786: gap of 100 bp
132787 141286: contig of 8500 bp in length
141287 141386: gap of 100 bp
141387 147754: contig of 6368 bp in length
147755 147854: gap of 100 bp
147855 153419: contig of 5359 bp in length
153420 153519: gap of 100 bp
153520 157078: contig of 3559 bp in length

```

```

* 157079 157178: gap of 100 bp
* 157179 162389: contig of 5211 bp in length
* 162390 162489: gap of 100 bp
* 162490 164934: contig of 2445 bp in length
* 164935 165034: gap of 100 bp
* 165035 168246: contig of 3212 bp in length
* 168247 168346: gap of 100 bp
* 168347 170645: contig of 2239 bp in length
* 170646 170745: gap of 100 bp
* 170746 172143: contig of 1398 bp in length
* 172144 172243: gap of 100 bp
* 172244 173516: contig of 1273 bp in length.

```

FEATURES

source

```

1..173516
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CTD-3243P17"

```

```

misc_feature
1..21771
/note="assembly_fragment"
21872..41497
/note="assembly_fragment"
41598..57282
/note="assembly_fragment"
57383..72819
/note="assembly_fragment"
72920..86720
/note="assembly_fragment"
86821..97172
/note="assembly_fragment"
97273..107979
/note="assembly_fragment"
108080..117581
/note="assembly_fragment"
117682..125644
/note="assembly_fragment"
125745..132686
/note="assembly_fragment"
132787..141286
/note="assembly_fragment"
141387..147754
/note="assembly_fragment"
147855..153419
/note="assembly_fragment"
153520..157078
/note="assembly_fragment"
157179..162389
/note="assembly_fragment"
162490..164934
/note="assembly_fragment"
165035..168246
/note="assembly_fragment"
168347..170645
/note="assembly_fragment"
170746..172143
/note="assembly_fragment"
172244..173516
/note="assembly_fragment"

```

ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 173516;
 Best Local Similarity 95.0%; Pred.No. 39;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCGATTCCTAGAGGAAGGA 20
 |||||
 Db 152267 GAGATTCCTAGAGGAAGGA 152286

RESULT 11
 AC025076/c

```

LOCUS      AC025076               183298 bp      DNA      linear      HTG 26-MAY-2000
DEFINITION Homo sapiens clone RP11-156K16, WORKING DRAFT SEQUENCE, 18
unordered pieces.
ACCESSION  AC025076
VERSION    AC025076.3  GI:8077051
KEYWORDS   HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 183298)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens, clone RP11-156K16
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 183298)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavskiy,L., Bouckhalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
            Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On May 25, 2000 this sequence version replaced gi:7272259.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www.seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L7828
            Center clone name: 156_K_16
            ----- Summary Statistics
            Sequencing vector: M13; M7815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 174315 bases at least Q40
            Consensus quality: 178612 bases at least Q30
            Consensus quality: 180475 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 181598; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.6 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 1774: contig of 1774 bp in length
            * 1775 1874: gap of 100 bp
            *
            1875 4363: contig of 2489 bp in length
            4364 4463: gap of 100 bp
            4464 7303: contig of 2840 bp in length
            7304 7403: gap of 100 bp
            7404 10043: contig of 2640 bp in length
            10044 10143: gap of 100 bp
            10144 14123: contig of 3980 bp in length
            14124 14223: gap of 100 bp
            14224 18507: contig of 4284 bp in length
            18508 21538: contig of 2931 bp in length
            21539 21638: gap of 100 bp
            21639 27268: contig of 5630 bp in length
            27269 33110: contig of 100 bp
            33111 33210: gap of 100 bp
            33211 39973: contig of 6763 bp in length
            39974 40073: gap of 100 bp
            40074 45638: contig of 5565 bp in length
            45639 45738: gap of 100 bp
            45739 55034: contig of 9296 bp in length
            55035 55134: gap of 100 bp
            55135 62645: contig of 7511 bp in length
            62646 62745: gap of 100 bp
            62746 77397: contig of 14652 bp in length
            77398 93454: contig of 15957 bp in length
            93455 110875: contig of 17321 bp in length
            110876 110975: gap of 100 bp
            110976 128888: contig of 17913 bp in length
            128889 182988: gap of 100 bp
            182989 183298: contig of 5430 bp in length.

FEATURES             Location/Qualifiers
     source            1..183298
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /clone_lib="RP11-156K16"
                        /clone_lib="RP11-156K16"
     misc_feature      1..1774
                        /note="assembly_fragment"
     misc_feature      1875..4363
                        /note="assembly_fragment"
     misc_feature      4464..7303
                        /note="assembly_fragment"
     misc_feature      7404..10043
                        /note="assembly_fragment"
     misc_feature      10144..14123
                        /note="assembly_fragment"
     misc_feature      14224..18507
                        /note="assembly_fragment"
     misc_feature      18508..21538
                        /note="assembly_fragment"
                        clone_end:T7
                        vector_side:left
     misc_feature      21639..27268
                        /note="assembly_fragment"
     misc_feature      27369..33110
                        /note="assembly_fragment"
     misc_feature      33211..39973
                        /note="assembly_fragment"
     misc_feature      40074..45638
                        /note="assembly_fragment"
     misc_feature      45739..55034
                        /note="assembly_fragment"
     misc_feature      55135..62645
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:right
     misc_feature      62746..77397
                        /note="assembly_fragment"
     misc_feature      77498..93454

```

```

/note="assembly_fragment"
93555..110875
/note="assembly_fragment"
110976..128888
/note="assembly_fragment"
128989..183298
/note="assembly_fragment"

ORIGIN
Query Match      87.6%; Score 18.4; DB 2; Length 183298;
Best Local Similarity 95.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCGATTCCTAGAGGAAGGA 20
Db 26670 GAGATTCCTAGAGGAAGGA 26651

RESULT 12
AP001541 212884 bp DNA linear PRI 15-MAR-2003
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-68304,
DEFINITION complete sequence.
ACCESSION AP001541
VERSION AP001541.4 GI:13422924
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA
Published Only in Database (2000)
2 (bases 1 to 212884)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (24-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suhiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@sc.riken.go.jp, URL:http://hgp.sc.riken.go.jp/,
Tel.81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Mar 21, 2001 this sequence version replaced gi:10130042.
FEATURES
source
location/Qualifiers
1..212884
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-68304"

ORIGIN
Query Match      87.6%; Score 18.4; DB 9; Length 212884;
Best Local Similarity 95.0%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCGATTCCTAGAGGAAGGA 20
Db 91524 GAGATTCCTAGAGGAAGGA 91543

RESULT 13
AC106477 245820 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-165F14, *** SEQUENCING IN PROGRESS
DEFINITION ***; 3 unordered pieces.
ACCESSION AC106477
VERSION AC106477.5 GI:30580899
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)

```

```

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 245820)
Muzny,D.,Marie,, Metsker,M.,Lee,, Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpaty,S., Kelly,S., Khan,Z., King,L., Kovar,C., Liu,J.,
Kows,C., Krat,C.L., Lebow,H., Levan,J., Lewis,L., Li-Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshew,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwankwelu,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzos,J., Quiroz,J., Rachlin,E., Reeves,K., Regier,M., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 245820)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 245820)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24819137.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.tmc.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas

```

```

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 245820)
Muzny,D.,Marie,, Metsker,M.,Lee,, Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpaty,S., Kelly,S., Khan,Z., King,L., Kovar,C., Liu,J.,
Kows,C., Krat,C.L., Lebow,H., Levan,J., Lewis,L., Li-Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshew,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwankwelu,O., Okwuonu,G., Olarnpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzos,J., Quiroz,J., Rachlin,E., Reeves,K., Regier,M., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 245820)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 245820)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24819137.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.tmc.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas

```

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCW

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GKZI

Center clone name: CH230-165F14

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 225710 bases at least Q40

Consensus quality: 229029 bases at least Q30

Consensus quality: 231433 bases at least Q20

Estimated insert size: 235116; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 242644: contig of 242644 bp in length

* 242645 242744: gap of unknown length

* 242745 243368: contig of 1624 bp in length

* 244369 244468: gap of unknown length

* 244469 245820: contig of 1352 bp in length.

----- Location/Qualifiers

1. 245820

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-165F14"

1. 1822

/note="wgs_end_extension"

clone_end:77"

2293..2621

/note="clone_boundary"

clone_end:77"

site:EcoRI

end_sequence:BH273380"

241929..242240

/note="clone_boundary"

clone_end:Sp6

site:EcoRI

end_sequence:BH273381"

ORIGIN

Query Match 87.6%; Score 19.4; DB 2; Length 245820;

Best Local Similarity 95.0%; Pred. No. 38;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGGATTCTAGAGGAAGGA 20

|||||

Db 26316 GGGATTCTAGAGGTAGGA 26335

|||||

RESULT 14

AC016931

LOCUS

DEFINITION Homo sapiens 3 BAC RP11-22E12 (Roswell Park Cancer Institute Human

BAC Library) complete sequence.

AC016931

VERSION

AC016931.21 GI:21617624

HTG.

Source

Homo sapiens (human)

ORGANISM

1 (bases 1 to 169032)

REFERENCE

AUTHORS

Muzny, D.M., Adams, C., Ali-Osman, F.R., Allen, C.,

Albrooks, S.L., Anaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,

Barbata, J., Benton, J., Binage, K., Blankenburg, K., Bonin, D.,

Bouay, J., Bowie, S., Brieau, M., Brown, E., Brown, M., Bryant, N.F.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorelli, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,

Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,

Hollings, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,

Louise, H., Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,

Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,

Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,

Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D.,

Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N.,

Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N.,

Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L., Peters, L.,

Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M.,

Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S.,

Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E.,

Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A.,

Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C.,

Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L.,

Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S.,

Warren, K., Washington, C., Watlington, S., Williams, G.,

Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.,

Wu, Y.F., Zhou, J., Zorrilla, S., Zorrilla, S.L., Weinstein, G. and

Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 169032)

Worley, K.C.

Direct Submission

Submitted (09-DEC-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 169032)

Worley, K.C.

Direct Submission

Submitted (23-JUN-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 169032)

Worley, K.C.

Direct Submission

Submitted (27-JUN-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 169032)

Worley, K.C.

Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187214)
Biren, B., Nussbaum, C. and Lander, E.
Mus musculus, clone RP23-1N19
Unpublished
2 (bases 1 to 187214)
Biren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Baetien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazarro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kanat, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
Matthews, C., McCarthy, M., Melnick, J., Meneus, L., Mihova, T.,
Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, F., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187214)
Biren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kanat, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Melnick, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Vankataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2003 this sequence version replaced gi:28195913.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26269
Center clone name: 1.N.19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 184963 bases at least Q40
Consensus quality: 185729 bases at least Q30
Consensus quality: 186054 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 186414; sum-of-contigs
Quality coverage: 11.3 in Q20 bases; agarose-fp

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Quality coverage: 10.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3471: contig of 3471 bp in length
* 3472 3571: gap of 100 bp
* 3572 4642: contig of 1071 bp in length
* 4643 4742: gap of 100 bp
* 4743 6413: contig of 1671 bp in length
* 6414 6513: gap of 100 bp
* 6514 11072: contig of 4559 bp in length
* 11073 11172: gap of 100 bp
* 11173 14809: contig of 3637 bp in length
* 14810 14909: gap of 100 bp
* 14910 39178: contig of 24269 bp in length
* 39179 39278: gap of 100 bp
* 39279 105215: contig of 65937 bp in length
* 105216 105315: gap of 100 bp
* 105316 154554: contig of 49339 bp in length
* 154555 154754: gap of 100 bp
* 154755 187214: contig of 32460 bp in length.

Location/Qualifiers
1..187214
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RP23-1N19"
/clone_lib="RP23-1N19"
1..3471
/note="assembly_fragment"
clone_end:SP6
vector_side:left
3572..4642
/note="assembly_fragment"
4743..6413
/note="assembly_fragment"
6514..11072
/note="assembly_fragment"
11173..14809
/note="assembly_fragment"
14910..39178
/note="assembly_fragment"
39279..105215
/note="assembly_fragment"
105316..154654
/note="assembly_fragment"
154755..187214
/note="assembly_fragment"
clone_end:T7
vector_side:right

ORIGIN
Query Match 85.7%; Score 18; DB 2; Length 187214;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGATTCCTAGAGGAAG 18
|||||||
Db 37785 GGGATTCCTAGAGGAAG 37768
Search completed: March 25, 2004, 12:49:59
Job time: 148.084 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 07:40:23 ; Search time 21.7931 Seconds

(without alignments)
4093.601 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423

Perfect score: 21

Sequence: 1 gggattcctagaggagag 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2000s:*
5: Geneseqn2000s:*
6: Geneseqn2000s:*
7: Geneseqn2000s:*
8: Geneseqn2000s:*
9: Geneseqn2000s:*
10: Geneseqn2000s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	6021	6	ABK48986 Genomic D
2	21	100.0	6458	6	ABK48984 Genomic D
3	19.4	92.4	8865	4	ABK87029 Human imm
4	17.8	84.8	9951	6	ABK39928 Human che
5	17.4	82.9	549	7	ABX73141 Metabolic
6	16.8	80.0	2796	3	AAA28713 C. pneumo
7	16.8	80.0	2817	6	ABL91186 Chlamydia
8	16.8	80.0	3000	3	AAA28712 C. pneumo
9	16.8	80.0	32177	4	AA529828 Human cyt
10	16.8	80.0	32177	4	AA105374 Human rep
11	16.8	80.0	32177	5	ABK18493 Human ner
12	16.8	80.0	110000	2	AA591990_05
13	16.8	80.0	273254	3	AA581914 Chlamydia
14	16.8	80.0	325791	4	AA543104 Human Oes
15	16.4	78.1	303	3	AA524739 Human sec
16	16.4	78.1	307	6	ABN60795 Human can
17	16.4	78.1	422	8	ACH21145 Human adu
18	16.4	78.1	436	6	ABV97368 Human pan
19	16.4	78.1	591	2	AA597629 Extended
20	16.4	78.1	610	9	AB555830 Toxicity-
21	16.4	78.1	1358	9	AA562425 cDNA sequ
22	16.4	78.1	2625	4	AAH14444 Human CDN
23	16.4	78.1	2805	6	ABL91187 Chlamydia

C 24	16.4	78.1	3003	3	AAA28711	Aaa28711 C. pneumo
C 25	16.4	78.1	3200	3	AAA28710	Aaa28710 C. pneumo
C 26	16.4	78.1	29956	8	ADA02963	Ada02963 Mouse Lck
C 27	16.4	78.1	29956	9	ADB72701	Adb72701 Mouse Lck
C 28	16.4	78.1	29956	9	ADC85443	Adc85443 Mouse Lck
C 29	16.4	78.1	31749	4	AAK72959	Aak72959 Human imm
C 30	16.4	78.1	78785	7	AAI60948	Aai60948 Human nep
C 31	16.4	78.1	78925	3	AAAC8988	Aac89888 Human FN
C 32	16.4	78.1	110000	7	ACF42745_1	Continuation (2 of
C 33	16.4	78.1	202001	6	ABS52506	ABs52506 Human tra
C 34	16.2	77.1	332	6	ABS72685	AbS72685 Human gen
C 35	16.2	77.1	384	6	ABN17597	Abn17597 Human ORF
C 36	16.2	77.1	479	8	ACH44328	Ach44328 Human foe
C 37	16.2	77.1	520	4	AAI16066	Aai16066 Human bre
C 38	16.2	77.1	600	5	ABV54015	ABv54015 Human pro
C 39	16.2	77.1	900	5	AA592506	Aa592506 Human T2R
C 40	16.2	77.1	900	5	AA570020	Aas70020 DNA encod
C 41	16.2	77.1	3117	5	ABA15179	AbA15179 Human ner
C 42	16.2	77.1	3117	5	ABA15177	AbA15177 Human ner
C 43	16.2	77.1	3117	5	ABA15180	AbA15180 Human ner
C 44	16.2	77.1	3117	5	ABA15178	AbA15178 Human ner
C 45	16.2	77.1	3172	6	AA562320	Aa562320 cDNA sequ

ALIGNMENTS

RESULT 1

ABK48986

ID ABK48986 standard; DNA; 6021 BP.

XX

AC ABK48986;

DT 02-JUL-2002 (first entry)

XX

DE Genomic DNA encoding mouse mesenchyme forkhead-1 (MHF-1)/FOXO2.

XX

KW Transcription factor; FOXO2; antidiabetic; anorectic; antilipaeamic;

KW cardiovascular; FOXO2 polypeptide expression modulator; FOXO2 promoter;

KW diabetes mellitus; obesity; hypercholesterolaemia; dyslipidaemia;

KW cardiovascular disease; mouse mesenchyme forkhead 1; MHF-1; gene; ds.

XX

OS Mus musculus.

XX

FH Key

FT CDS

FT Location/Qualifiers

FT 2070..3554

FT /*tag= a

FT /product= "MHF-1 or FOXO2"

FT /note= "Mouse mesenchyme forkhead 1"

XX

XX WO200227008-A1.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC The invention describes an isolated human FOXC2 promoter region. The
CC mammalian FOXC2 promoter region for the modulation of FOXC2 polypeptide
CC expression has therapeutic value in treating type II diabetes mellitus,
CC obesity, hypercholesterolaemia, other cardiovascular diseases or
CC dyslipidaemias. This sequence encodes the mouse mesenchyme forkhead 1
CC (MHF-1) protein (also called FOXC2 transcription factor), described in
CC the method of the invention
XX
SQ Sequence 6021 BP; 1350 A; 1675 C; 1642 G; 1354 T; 0 U; 0 Other;
Query Match 100.0%; Score 21; DB 6; Length 6021;
Best Local Similarity 100.0%; Pred. No. 2.5; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGATTCCTAGAGGAGGAG 21
|||||
Db 385 GGGATTCCTAGAGGAGGAG 405
RESULT 2
ID ABK48984 standard; DNA; 6458 BP.
XX
AC ABK48984;
XX
DT 02-JUL-2002 (first entry)
XX
DE Genomic DNA encoding human transcription factor FOXC2.
XX
KW Transcription factor; FOXC2; antidiabetic; anorectic; antilipaemic;
KW cardiovascular; FOXC2 polypeptide expression modulator; FOXC2 promoter;
KW diabetes mellitus; obesity; hypercholesterolaemia; dyslipidaemia;
KW cardiovascular disease; gens; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 1..215
FT /tag= b
FT /note= "First exon according to the alternative
FT transcript"
FT 5'UTR 1..186
FT /tag= a
FT misc_feature 187..215
FT /tag= c
FT /note= "Region coding for 5' part of alternative protein"
FT misc_signal 215..216
FT /tag= d
FT /note= "Alternative first exon splice site"
FT misc_difference 216..475
FT /tag= e
FT /note= "Fragment of the FOXC2 enhancer. Specifically
FT claimed in claim 16"
FT misc_difference 223..231
FT /tag= f
FT /note= "Fragment of the FOXC2 enhancer. Specifically
FT claimed in claim 12"
FT misc_difference 359..375
FT /tag= g
FT /note= "Fragment of the FOXC2 enhancer. Specifically
FT claimed in claim 13"
FT misc_difference 375..402
FT /tag= h
FT /note= "Fragment of the FOXC2 enhancer. Specifically
FT claimed in claim 14"
FT misc_difference 403..423
FT /tag= i
FT /note= "Fragment of the FOXC2 enhancer. Specifically
FT claimed in claim 15"
FT misc_difference 1250..2235
FT /tag= k
FT /note= "Fragment of the FOXC2 promoter. Specifically
FT claimed in claim 3"

FT misc_difference 1250..1749
FT /tag= i
FT /note= "Fragment of the FOXC2 promoter. Specifically
FT claimed in claim 2"
FT misc_difference 1692..1703
FT /tag= l
FT /note= "Fragment of the FOXC2 promoter. Specifically
FT claimed in claim 1"
FT exon 1745..4629
FT /tag= n
FT /note= "First exon according to the published form of the
FT transcript"
FT 5'UTR 1746..2234
FT /tag= m
FT CDS 2235..3740
FT /tag= o
FT /product= "FOXC2"
FT /note= "Transcription factor"
FT misc_binding 2448..2735
FT /tag= p
FT /note= "Region coding for DNA-binding domain"
FT exon 2516..4629
FT /tag= r
FT /note= "Second exon according to the alternative
FT transcript"
FT misc_feature 2516..3740
FT /tag= q
FT /note= "Portion of polypeptide used in alternative
FT transcript"
FT 3'UTR 3741..4629
FT /tag= s
XX WC200227008-A1.
XX
XX 04-APR-2002.
XX
XX 26-SEP-2001; 2001WO-SE002098.
XX
XX 26-SEP-2000; 2000SE-00003435.
PR 10-OCT-2000; 2000US-0238897P.
PR 09-NOV-2000; 2000SE-00004102.
XX
XX (BIOV-) BIOVITRUM AB.
XX
XX Enerbaeck S, Krook K, Rondahl L, Wasserman WW;
XX WPI; 2002-352129/39.
DR P-PSDB; AAU79916.
XX
XX An isolated FOXC2 promoter region that modulates the expression of a
XX FOXC2 polypeptide is useful for treating obesity and type II diabetes
XX mellitus.
XX
XX Claim 3; Page 33-38; 62pp; English.
XX
XX The invention describes an isolated human FOXC2 promoter region. The
XX mammalian FOXC2 promoter region for the modulation of FOXC2 polypeptide
XX expression has therapeutic value in treating type II diabetes mellitus,
XX obesity, hypercholesterolaemia, other cardiovascular diseases or
XX dyslipidaemias. This sequence encodes the human transcription factor
XX FOXC2, the promoter of which is described in the invention
XX
SQ Sequence 6458 BP; 1443 A; 1826 C; 1715 G; 1474 T; 0 U; 0 Other;
Query Match 100.0%; Score 21; DB 6; Length 6458;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGATTCCTAGAGGAGGAG 21
|||||
Db 403 GGGATTCCTAGAGGAGGAG 423

RESULT 3
AAK87029
ID AAK87029 standard; DNA; 8865 BP.
XX
AC AAK87029;
XX
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41841.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217486P.
XX 14-JUL-2000; 2000US-0218293P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226682P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231124P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240360P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.

PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 41841; 3071pp + Sequence Listing; English.
XX AA054951 to AA064702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AA082170 to AA082191. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AA064703
CC to AA087694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AA054942 to AA054950 and AA082169
CC represent sequences used in the exemplification of the present invention
XX
XX SQ Sequence 8865 BP; 2486 A; 1904 C; 2073 G; 2401 T; 0 U; 1 Other;
Query Match 92.4%; Score 19.4; DB 4; Length 8865;
Best Local Similarity 95.2%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GGGATTCTCAGGAGGAGGAG 21
Db 857 GGGATTCTCAGGAGGAGGAG 877
RESULT 4
ABK39928
ID ABK39928 standard; DNA; 9951 BP.
XX AC ABK39928;
XX DT 21-MAY-2002 (first entry)
XX DE Human chemically pretreated gene sequence #5 strand 1.
XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytosstatic; ALDH6; CYP11A; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX Homo sapiens.
XX WO200202806-A2.
XX 10-JAN-2002.
XX 29-JUN-2001; 2001WO-EP007470.

XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154757/20.
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
XX pharmacogenomics and for therapy of diseases e.g. cancer.
XX Claim 1; SEQ ID NO 9; 24pp; English.
XX The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001975), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004956),
CC NM_019300, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 9951 BP; 2117 A; 204 C; 2950 G; 4680 T; 0 U; 0 Other;
Query Match 84.8%; Score 17.8; DB 6; Length 9951;
Best Local Similarity 90.5%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GGGATTCTCAGGAGGAGGAG 21
Db 9036 GGGATTCTCAGGAGGAGGAG 9056
RESULT 5
ABX73141
ID ABX73141 standard; cDNA; 549 BP.
XX AC ABX73141;
XX DT 17-MAR-2003 (first entry)
XX DE Metabolic pathway (MP) protein cDNA #23.
XX Metabolic pathway protein; MP; gene; ss; moss; alga; Ceratodon purpureus;
KW fine chemical; Corynebacterium; Brevibacterium; vitamin; cofactor;
KW metabolism; essential amino acid; tocopherol; riboflavin; vitamin C.
XX Physcomitrella patens.
XX OS
XX US2002142422-A1.
XX 03-OCT-2002.
XX 12-DEC-2000; 2000US-00734017.

XX PR 16-DEC-1999; 99US-0171100P.
XX PA (LERC/) LERCHL J.
XX PA (RENZ/) RENZ A.
XX PA (EHRH/) EHRHARDT T.
XX PA (REIN/) REINDL A.
XX PA (CIRP/) CIRPUS F.
XX PA (BISC/) BISCHOFF F.
XX PA (FRAN/) FRANK M.
XX PA (DUEW/) DUWENIG E.
XX PA (SCHM/) SCHMIDT R.
XX PA (RESK/) RESKI R.
XX PI Lersch J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;
XX PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;
XX PR WPI: 2003-155946/15.
XX PR P-PSDB; ABUS4893.
XX PR Novel isolated nucleic acids from moss e.g., Physcomitrella patens,
XX PT encoding a metabolic pathway protein, useful for modifying production of
XX PT amino acids, vitamins, cofactors in plants, algae and microorganisms.
XX PS Disclosure; Page 36; 107pp; English.
XX CC The invention relates to a nucleic acid molecule from a moss (e.g.,
XX CC Physcomitrella patens or Ceratodon purpureus) encoding a metabolic acid
XX CC pathway (MP) protein or its portion. A vector comprising a nucleic acid
XX CC is useful for producing a fine chemical which involves culturing a cell
XX CC containing the vector such that the fine chemical is produced and then
XX CC recovered. The cell is a microorganism belonging to the genus
XX CC Corynebacterium or Brevibacterium, to the genus mosses or algae, or is a
XX CC plant cell. The expression of the nucleic acid from the vector in the
XX CC cell results in modulation of the production of the fine chemical such as
XX CC amino acids, vitamins, cofactors, nucleotides and/or nucleosides. The MP
XX CC nucleic acid molecules are useful for modulating production of fine
XX CC chemicals in microorganisms, algae and plants, either directly or
XX CC indirectly. Plant genes originating from P. patens are useful for
XX CC modifying metabolism of essential amino acids, tocopherol, riboflavin,
XX CC vitamin C, etc., in plants as well as algae and microorganisms, enabling
XX CC these host cells to increase their capacity to produce the respective
XX CC components as well as improving survival and fitness of the cell. The
XX CC nucleic acids are useful for identifying an organism as being P. patens
XX CC or its close relative in a mixed population of microorganisms, as markers
XX CC for specific regions of the genome and for evolutionary and protein
XX CC structural studies. Sequences ABX73119-ABX73160 represent cDNA molecules
XX CC encoding MP proteins of the invention
XX SQ Sequence 549 BP; 120 A; 136 C; 148 G; 145 T; 0 U; 0 Other;
Query Match 82.9%; Score 17.4; DB 7; Length 549;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCGATTCCTAGAGGGAAGG 19
Db 247 GCGATTCCTAGAGGGAAGG 265
RESULT 6
AAA28713/c
ID AAA28713 standard; DNA; 2796 BP.
XX AC AAA28713;
XX AC AAA28713;
XX DT 15-SEP-2003 (revised)
XX DT 29-AUG-2000 (first entry)
XX DE C. pneumoniae CPN:00628 open reading frame.
XX DE Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KW Artgen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;

KW anti-arteriosclerotic; vaccine; ds.
XX OS Chlamydophila pneumoniae.
XX PN WO200024765-A2.
XX PD 04-MAY-2000.
XX PF 28-OCT-1999; 99WO-CA000992.
XX PR 28-OCT-1998; 98US-0106034P.
XX PR 28-OCT-1998; 98US-0106039P.
XX PR 28-OCT-1998; 98US-0106042P.
XX PR 28-OCT-1998; 98US-0106044P.
XX PR 29-OCT-1998; 98US-0106072P.
XX PR 29-OCT-1998; 98US-0106073P.
XX PR 29-OCT-1998; 98US-0106074P.
XX PR 29-OCT-1998; 98US-0106087P.
XX PR 02-NOV-1998; 98US-0106587P.
XX PR 02-NOV-1998; 98US-0106588P.
XX PR 02-NOV-1998; 98US-0106589P.
XX PR 02-NOV-1998; 98US-0107034P.
XX PR 02-NOV-1998; 98US-0107035P.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Murdin AD, Oomen RP, Wang J;
XX PI WPI: 2000-350688/30.
XX PR P-PSDB; ANY92833.
XX CC Chlamydia antigens and the proteins they encode, useful for vaccinating
XX PT against Chlamydia infections that affect the respiratory tract.
XX PT Claim 2; Fig 23; 226pp; English.
XX CC The nucleic acids may be used for the recombinant production of the
XX CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
XX CC recombinant DNA methodologies. The polypeptides may then be used to
XX CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as C.
XX CC pneumoniae, are pathogens responsible for upper respiratory tract
XX CC infections such as community acquired pneumonia, acute respiratory
XX CC disease and bronchitis and may be implicated in atherosclerotic changes
XX CC and asthma. The nucleic acids may also be used as probes for detecting
XX CC the presence of Chlamydia nucleic acids in samples (and therefore
XX CC diagnose infections) and the proteins may be used as antigens for the
XX CC production of antibodies that may be used to detect Chlamydia proteins in
XX CC samples (e.g. via enzyme linked immunosorbant assay (ELISA)). (Updated on
XX CC 15-SEP-2003 to standardise OS field)
XX SQ Sequence 2796 BP; 798 A; 703 C; 528 G; 767 T; 0 U; 0 Other;
Query Match 80.0%; Score 16.8; DB 3; Length 2796;
Best Local Similarity 90.0%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GCGATTCCTAGAGGGAAGG 21
Db 2509 GCGATTCCTAGAGGGAAGG 2490
RESULT 7
ABL91186/c
ID ABL91186 standard; DNA; 2817 BP.
XX AC ABL91186;
XX AC ABL91186;
XX DT 29-AUG-2003 (revised)
XX DT 29-JUL-2002 (first entry)
XX DE Chlamydia pneumoniae cp6751 ORF DNA, SEQ ID NO:6.
XX DE Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
KW Chlamydial infection; antigen; immunogen; vaccine; diagnosis;

human respiratory disease; cardiovascular disease; atherosclerosis;
coronary artery disease; carotid artery stenosis; myocardial infarction;
cerebrovascular disease; aortic aneurysm; claudication; stroke;
strain CML029; open reading frame; ORF; Gene; ds.

Chlamydomonada pneumoniae.

Key Location/Qualifiers
CDS 1..2817
/tag= a
/product= "cp6751"
FT sig_peptide 1..51
/tag= b
FT mat_peptide 52..2814
/tag= c
/product= "Mature protein"

WO200202606-A2.

10-JAN-2002.

03-JUL-2001; 2001WO-IB001445.

03-JUL-2000; 2000GB-00016363.

11-JUL-2000; 2000GB-00017047.

21-JUL-2000; 2000GB-00017983.

07-AUG-2000; 2000GB-00019368.

18-AUG-2000; 2000GB-00020440.

14-SEP-2000; 2000GB-00022583.

10-NOV-2000; 2000GB-00027549.

22-DEC-2000; 2000GB-00031706.

(CHIR-) CHIRON SPA.

Ratti G, Grandi G;

WPI: 2002-154726/20.

N-PSDB; ABB90528.

Novel Chlamydia pneumoniae protein useful in the manufacture of a
medicament for treatment or prevention of infection due to Chlamydia,
preferably Chlamydia pneumoniae, and for diagnostic purposes.

Claim 5; Page 43-44; 364pp; English.

Sequences ABB30526-ABB90715 represent novel proteins from Chlamydia
pneumoniae (strain CML029), and ABL91184-ABL91373 represent DNA encoding
them. The proteins are predicted to be immunogenic and may therefore be
useful in vaccine production and for diagnostic purposes. Chlamydia
pneumoniae is a common cause of respiratory disease in humans, and is
also involved in the development of cardiovascular diseases such as
atherosclerosis, coronary artery disease, carotid artery stenosis,
myocardial infarction, cerebrovascular disease, aortic aneurysm,
claudication and stroke. The proteins and nucleic acids of the invention
may be used in vaccines and pharmaceutical compositions for the
prevention or treatment of Chlamydia infections, particularly Chlamydia
pneumoniae infections. The proteins may also be used in the detection of
Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
DNA probe assay or blotting techniques for determining Chlamydia
pneumoniae gene expression. The present sequence represents a
specifically claimed DNA which encodes a Chlamydia pneumoniae protein of
the invention. (Updated on 29-AUG-2003 to standardise OS field)

Sequence 2817 BP; 800 A; 706 C; 534 G; 777 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 6; Length 2817;

Best Local Similarity 90.0%; Pred. No. 2.5e-02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 GGATTCCTAGAGGGAAGGAG 21

|||||

2530 GGATTCCTACAGGAGGAG 2511

|||||

RESULT 8

AAA28712/c

ID AAA28712 standard; DNA; 3000 BP.

XX AC AAA28712;

XX 15-SEP-2003 (revised)

DT 29-AUG-2000 (first entry)

XX C. pneumoniae CPN100628 gene.

XX Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;

KW anti-arteriosclerotic; vaccine; ds.

XX Chlamydomonada pneumoniae.

XX Key Location/Qualifiers

FT CDS 101..2896

FT /tag= a

XX WO200024765-A2.

XX 04-MAY-2000.

XX 28-OCT-1999; 99WO-CA000992.

XX 28-OCT-1998; 98US-0106034P.

XX 28-OCT-1998; 98US-0106039P.

XX 28-OCT-1998; 98US-0106042P.

XX 28-OCT-1998; 98US-0106044P.

XX 29-OCT-1998; 98US-0106072P.

XX 29-OCT-1998; 98US-0106073P.

XX 29-OCT-1998; 98US-0106074P.

XX 29-OCT-1998; 98US-0106087P.

XX 02-NOV-1998; 98US-0106587P.

XX 02-NOV-1998; 98US-0106588P.

XX 02-NOV-1998; 98US-0107034P.

XX 02-NOV-1998; 98US-0107035P.

XX (CONN-) CONNAUGHT LAB LTD.

XX Murdin AD, Comen RP, Wang J;

XX WPI: 2000-350688/30.

XX P-PSDB; AAY92833.

Chlamydia antigens and the proteins they encode, useful for vaccinating
against Chlamydia infections that affect the respiratory tract.
Claim 2; Fig 23; 226pp; English.
The nucleic acids may be used for the recombinant production of the
Chlamydia polypeptides (either in vivo or in vitro) according to standard
recombinant DNA methodologies. The polypeptides may then be used to
vaccinate against Chlamydia infections in mammals. Chlamydia, such as C.
pneumoniae, are pathogens responsible for upper respiratory tract
infections such as community acquired pneumonia, acute respiratory
disease and bronchitis and may be implicated in atherosclerotic changes
and asthma. The nucleic acids may also be used as probes for detecting
the presence of Chlamydia nucleic acids in samples (and therefore
diagnose infections) and the proteins may be used as antigens for the
production of antibodies that may be used to detect Chlamydia proteins in
samples (e.g. via enzyme linked immunosorbent assay (ELISA)). (Updated on
15-SEP-2003 to standardise OS field)

Sequence 3000 BP; 865 A; 739 C; 562 G; 834 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 3; Length 3000;

Best Local Similarity 90.0%; Pred. No. 2.5e-02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GGATTCCTAGAGGAGGAG 21
Db 2609 GGATTCCTACAGGAGGAG 2590

RESULT 9
AAS29828/c
ID AAS29828 standard; DNA; 32177 BP.
XX AC AAS29828;
XX DT 21-NOV-2001 (first entry)
XX DE Human cytoskeletal element-related polypeptide encoding genomic DNA #15.
XX KW Cytoskeletal element-related protein; human; mouse; rabbit; goat; horse;
KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; neurotropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility; food additive.
XX OS Homo sapiens.
XX PN WO200155168-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001331.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184684P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209457P.
PR 28-JUN-2000; 2000US-0214986P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220953P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225256P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234597P.
PR 25-SEP-2000; 2000US-0234598P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.

PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249224P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250180P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251888P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-476182/51.
DR
XX Novel isolated human cytoskeletal element-related polypeptide useful for
PT diagnosis/treatment of neoplastic disorders, disorders associated with
PT neural transmission, chromosomal abnormalities, autoimmune disorders.
XX
XX Claim 1; SEQ ID NO 113; 505pp; English.

XX Sequences AAS29814-AAS29838 represent genomic DNA molecules, which encode
CC the cytoskeletal element-related polypeptides of the invention.
CC Cytoskeletal polypeptides and their associated polynucleotides are useful
CC in the diagnosis, treatment and prevention of various types of disorders
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC sheep. A pathological condition can be determined by determining the
CC presence or absence of a mutation in a cytoskeletal polynucleotide. The
CC treatable disorders include autoimmune diseases such as rheumatoid
CC arthritis, hyperproliferative disorders such as neoplasms of the breast
CC or liver, cardiovascular disorders such as cardiac arrest,
CC cerebrovascular disorders such as cerebral ischaemia, nervous system
CC disorders such as Alzheimer's disease, infections caused by bacteria,
CC viruses and fungi, ocular disorders such as corneal infection, endocrine
CC disorders such as premature labour and infertility, gastrointestinal
CC disorders such as Crohn's disease, renal disorders such as
CC glomerulonephritis and respiratory disorders such as asthma. The
CC polypeptides can also be used to aid wound healing, to prevent skin aging
CC due to sunburn, to maintain organs before transplantation, to regenerate
CC tissues and in chemotaxis. The polypeptides can also be used as a food
CC additive or preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

Query Match 80.0%; Score 16.8; DB 4; Length 32177;
Best Local Similarity 90.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAGGA 20

Db 1139 GAGATCTCTAGAGGAGGA 1120

RESULT 10
AAL05374/c

ID AAL05374 standard; DNA; 32177 BP.
XX
AC AAL05374;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 8062.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
XX Homo sapiens.
OS
XX WO200155320-A2.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US0011339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0191213P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0215647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.

```
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234938P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246417P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246536P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249254P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 8062; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX SQ Sequence 32177 BP; 8826 A; 6801 C; 7130 G; 9420 T; 0 U; 0 Other;
XX
XX Query Watch 80.0%; Score 16.8; DB 4; Length 32177;
XX Best Local Similarity 90.0%; Pred. No. 2.9e+02;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GGGATTCTCTAGAGGAAGGA 20
XX | | | | | | | | | | | | | | | | | |
XX Db 1139 GAGATCTCTAGAGGAAGGA 1120

RESULT 11
ABAI8493/c
ID ABAI8493 standard; DNA; 32177 BP.
XX
XX AC ABAI8493;
XX
XX DT 23-JAN-2002 (first entry)
XX
XX DE Human nervous system related polynucleotide SEQ ID NO 10824.
XX
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200159063-A2.
XX
XX PD 16-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US001334.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
```

PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-02151135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225470P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226661P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227093P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241828P.
PR 01-NOV-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 10824; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AB11004-AB21534) and proteins
CC (AB114678-AB318001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and

CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 32177 BP; 8826 A; 6801 C; 7130 G; 9420 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 5; Length 32177;
Best Local Similarity 90.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAAGGA 20
DB 1139 GAGAACTCTAGAGGAAGGA 1120

RESULT 12

AAK91990_05/c
Continuation (6 of 13) of AAK91990 from base 500001 (Nucleotide sequence of the complete
WP Sequence split into 13 fragments LOCUS AAK91990 Accession Aak91990

WP	Fragment Name	Begin	End
WP	AAK91990_00	1	110000
WP	AAK91990_01	100001	210000
WP	AAK91990_02	200001	310000
WP	AAK91990_03	300001	410000
WP	AAK91990_04	400001	510000
WP	AAK91990_05	500001	610000
WP	AAK91990_06	600001	710000
WP	AAK91990_07	700001	810000
WP	AAK91990_08	800001	910000
WP	AAK91990_09	900001	1010000
WP	AAK91990_10	1000001	1110000
WP	AAK91990_11	1100001	1210000
WP	AAK91990_12	1200001	1230025

Query Match 80.0%; Score 16.8; DB 2; Length 110000;
Best Local Similarity 90.0%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGGATTCCTAGAGGAAGGAG 21
DB 46344 GGGATTCCTAGAGGAAGGAG 46325

RESULT 13

AAK81914
ID AAC81914 standard; DNA; 273254 BP.

AC AAK81914;

DT 15-SEP-2003 (revised)

DT 27-FEB-2001 (first entry)

XX Chlamydia pneumoniae genome DNA.

XX Genome; diagnosis; vaccine; ds.

XX Chlamydia pneumoniae.

XX WO200027994-A2.

XX 18-MAY-2000.

XX 12-NOV-1999; 99WO-US026923.

XX 12-NOV-1998; 98US-0108279P.

PR 08-APR-1999; 99US-0128606P.

XX (REGC) UNIV CALIFORNIA.

XX Stephens R, Mitchell W, Kalman S, Davis R;

XX WPI; 2000-376516/32.

XX Isolated nucleic acid for use in diagnostic and analytical methods
PT encodes genomic sequence of Chlamydia pneumoniae.

XX PS Claim 2; Page 128-320; 320pp; English.

XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (P1), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a
CC transcriptional initiation region functional in an expression host, and a
CC transcriptional termination region; (4) a cell comprising an expression
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell as a result of induction of the expression
CC cassette into the host cell, and the cellular progeny of the host cell;
CC (5) a method for producing a P1 comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other proteins
CC ; (6) a purified polypeptide composition comprising at least 50 weight %
CC of P1; and (7) a monoclonal antibody binding specifically to the peptide
CC of (6). (Updated on 15-SEP-2003 to standardise OS field)

XX SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 3; Length 273254;
Best Local Similarity 90.0%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGATTCCTAGAGGAAGGAG 21

DB 70053 GGATTCCTAGAGGAAGGAG 70072

RESULT 14

AAK431104
ID AAK431104 standard; DNA; 325791 BP.

AC AAK431104;

DT 18-DEC-2001 (first entry)

XX Human Oestrogen receptor beta gene.

XX Human; Oestrogen receptor beta; ERbeta; db; SNP; chromosome 6q.25.1;

XX single nucleotide polymorphism; cardiovascular disease;

XX autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;

XX osteoarthritis; osteoporosis; breast cancer; endometrial cancer.

XX Homo sapiens.

XX WO200162793-A2.

XX 30-AUG-2001.

XX 20-FEB-2001; 2001WO-US005360.

XX 22-FEB-2000; 2000US-0183755P.

XX 24-JAN-2001; 2001US-00768185.

XX (PEKE) PE CORP NY.

XX Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;

XX WPI; 2001-582041/65.
DR P-PSDB; AAU27322.
XX
PT Estrogen receptor gene and protein polymorphisms useful for diagnosis of
PT individuals at risk of developing bone disorders.
XX
XX Example 2; Fig 1; 245pp; English.
XX
XX The invention relates to a novel isolated peptide comprising or
CC consisting of an amino acid sequence selected from an amino acid sequence
CC of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC 10 amino acids), antibodies against them, nucleic acids encoding them
CC (including vectors for transforming cells). The gene for human ERbeta is
CC located on chromosome 6q.25.1. The variants are encoded by single
CC nucleotide polymorphisms (SNP). The variant peptides and proteins can be
CC used in assays to determine the biological activity of the protein, to
CC raise antibodies, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, to identify
CC compounds that modulate receptor activity and to screen compounds for the
CC ability to stimulate or inhibit interaction between the receptor protein
CC and a target molecule that normally interacts with the receptor protein
CC e.g. oestrogen. The antibody can be used to isolate the protein, to
CC assess expression in disease states e.g. cardiovascular disease and
CC autoimmune disease (e.g. systemic lupus erythematosus, arthritis,
CC rheumatism and osteoarthritis), osteoporosis, breast cancer and
CC endometrial cancer. In addition the antibodies can be used in
CC pharmacogenomic analysis and inhibiting protein function, e.g. blocking
CC the binding of the oestrogen receptor protein to a binding partner such
CC as a ligand. The nucleic acids encoding the proteins can be used as
CC probes, primers, chemical intermediates and in biological assays. The
CC present sequence is the human ERbeta gene
XX
SQ Sequence 325791 BP; 94098 A; 68292 C; 67970 G; 95431 T; 0 U; 0 Other;
Query Match 80.0%; Score 16.8; DB 4; Length 325791;
Best Local Similarity 90.0%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CGATTCTCTAGAGGGAGGAG 21
DB 263062 CGATTCTCTAGAGGGAGGAG 263081
RESULT 15
AAC24739/C
ID AAC24739 standard; cDNA; 303 BP.
XX
AC AAC24739;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 28814.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
PD
XX 21-FEB-2000; 2000EP-00200610.
PF
XX 26-FEB-1999; 95US-0122487P.
PR
XX (GSET) GENSET.
PA
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX
XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 28814; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 303 BP; 86 A; 66 C; 44 G; 99 T; 0 U; 8 Other;
Query Match 78.1%; Score 16.4; DB 3; Length 303;
Best Local Similarity 94.4%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 ATTCTCTAGAGGGAGGAG 21
DB 177 ATTCTCTAGAGGGAGGAG 160

Search completed: March 25, 2004, 10:25:24
Job time : 25.7931 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 09:55:14 ; Search time 4.13061 Seconds
(without alignments)
2821.370 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423

Perfect score: 21

Sequence: 1 gggattcctagagggaaggag 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
C 1	80.0	16.8	1230025	4	US-09-198-452A-1
C 2	78.1	16.4	202001	4	US-09-734-674-3
C 3	77.1	16.2	900	4	US-09-393-634-44
C 4	77.1	16.2	38653	4	US-09-922-445-1
C 5	75.2	15.8	287	4	US-09-313-294A-1436
C 6	75.2	15.8	955	4	US-09-641-638-13
C 7	75.2	15.8	955	4	US-09-641-638-14
C 8	75.2	15.8	43069	4	US-09-292-542A-1
C 9	73.3	15.4	3502	2	US-08-724-394A-16
C 10	72.4	15.2	288	4	US-09-621-976-12965
C 11	72.4	15.2	351	4	US-09-621-976-10679
C 12	72.4	15.2	463	4	US-09-621-976-2535
C 13	72.4	15.2	554	4	US-09-621-976-18505
C 14	72.4	15.2	642	4	US-09-252-991A-13569
C 15	72.4	15.2	810	4	US-09-252-991A-12999
C 16	72.4	15.2	873	4	US-09-252-991A-12472
C 17	72.4	15.2	893	3	US-09-313-300-5
C 18	72.4	15.2	1001	4	US-09-671-317-149
C 19	72.4	15.2	1686	4	US-09-833-381-2019
C 20	72.4	15.2	1707	4	US-09-596-248D-22
C 21	72.4	15.2	1766	4	US-09-517-467B-18
C 22	72.4	15.2	1787	3	US-08-923-454A-3
C 23	72.4	15.2	2526	1	US-07-912-952-1
C 24	72.4	15.2	2551	3	US-08-923-454A-26
C 25	72.4	15.2	3000	4	US-09-705-267A-18
C 26	72.4	15.2	3033	4	US-09-833-381-1810
C 27	72.4	15.2	3129	4	US-09-252-991A-13873

28 15.2 72.4 4131 4 US-09-252-991A-13773 Sequence 13773, A
 C 29 15.2 72.4 8211 4 US-09-252-991A-13656 Sequence 13656, A
 C 30 15.2 72.4 12425 4 US-09-616-289-50 Sequence 50, Appl
 C 31 15.2 72.4 15652 4 US-09-422-936-60 Sequence 60, Appl
 C 32 15.2 72.4 18000 4 US-09-657-346A-17 Sequence 17, Appl
 C 33 15.2 72.4 29604 3 US-08-781-891-207 Sequence 207, Appl
 C 34 15.2 72.4 29604 4 US-09-618-166-207 Sequence 207, Appl
 C 35 15.2 72.4 87350 3 US-08-781-891-79 Sequence 79, Appl
 C 36 15.2 72.4 87350 4 US-09-618-166-79 Sequence 79, Appl
 C 37 15.2 72.4 87543 4 US-09-791-211-3 Sequence 3, Appl
 C 38 15.2 72.4 161652 4 US-09-497-855A-40 Sequence 40, Appl
 C 39 15.2 72.4 169998 4 US-09-676-610B-24 Sequence 24, Appl
 C 40 15.2 72.4 197496 4 US-09-877-177A-10 Sequence 10, Appl
 C 41 14.8 70.5 512 4 US-09-621-976-3594 Sequence 3594, Ap
 C 42 14.8 70.5 1376 4 US-09-443-184-44 Sequence 44, Appl
 C 43 14.8 70.5 1390 4 US-09-205-258-124 Sequence 124, App
 C 44 14.8 70.5 2935 4 US-09-976-594-365 Sequence 365, App
 C 45 14.8 70.5 2935 4 US-09-976-594-365 Sequence 365, App

ALIGNMENTS

RESULT 1
 US-09-198-452A-1/c
 ; Sequence 1, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Griflais, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09198,452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 1
 ; LENGTH: 1230025
 ; TYPE: DNA
 ; ORGANISM: Chlamydia pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)-(15000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (15001)..(30000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (30001)..(45000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (45001)..(60000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (60001)..(75000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (75001)..(90000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (90001)..(105000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (105001)..(120000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (120001)..(135000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (135001)..(150000)
 ; OTHER INFORMATION: n=a or c or g or t
 ; NAME/KEY: misc feature
 ; LOCATION: (150001)..(165000)


```
; LOCATION: (8850001)..(900000)
; OTHER INFORMATION: n-a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (9000001)..(915000)
; OTHER INFORMATION: n-a or c or g or t
; NAME/KEY: misc_feature

Query Match      80.0%; Score 16.8; DB 4; Length 1230025;
Best Local Similarity 90.0%; Pred. No. 56;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGATTCCTAGAGGGAAGGAG 21
Db 546344 GGATTCCTACAGGAGGAG 546325

RESULT 2
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(202001)
; OTHER INFORMATION: n = A, T, C or G
;
US-09-734-674-3

Query Match      78.1%; Score 16.4; DB 4; Length 202001;
Best Local Similarity 94.4%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATTCCTAGAGGGAAGGAG 21
Db 124677 ATTCCTAGAGGGAAGGAG 124660

RESULT 3
US-09-393-634-44.
; Sequence 44, Application US/09393634
; Patent No. 6558910
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SP, a No. 6558910el Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; OTHER INFORMATION: human GR05
;
US-09-393-634-44

Query Match      77.1%; Score 16.2; DB 4; Length 900;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGGAAGGAG 21
Db 823 GGGATTCCTAGGGAAGGAG 843

RESULT 4
US-09-922-445-1/c
; Sequence 1, Application US/09922445
; Patent No. 6528268
; GENERAL INFORMATION:
; APPLICANT: Andersson, Maria K.
; APPLICANT: Berglund, Lars G. T.
; APPLICANT: Reneland, Rikard H.
; APPLICANT: Adam, Gail I. R.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR DETECTION OF HEART FAILURE
; FILE REFERENCE: GGI26US
; CURRENT APPLICATION NUMBER: US/09/922,445
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 38653
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(26156)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (24801)..(24801)
; OTHER INFORMATION: nucleotide 24801 is a single nucleotide polymorphism which can be
; OTHER INFORMATION: A or G
; NAME/KEY: misc feature
; LOCATION: (24941)..(24941)
; OTHER INFORMATION: nucleotide 24941 is a single nucleotide polymorphism which can be
; OTHER INFORMATION: T or C
; NAME/KEY: exon
; LOCATION: (26157)..(26252)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (26253)..(26401)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (26402)..(26543)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (26544)..(27024)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (27025)..(27178)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (27179)..(30519)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (27645)..(27645)
; OTHER INFORMATION: nucleotide 27645 is a single nucleotide polymorphism which can be
; OTHER INFORMATION: C or G
; NAME/KEY: exon
; LOCATION: (30520)..(30681)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (30682)..(30894)
; OTHER INFORMATION:
; NAME/KEY: exon
```


LOCATION: (30895)..(31027)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (31028)..(31747)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (31748)..(31841)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (31842)..(32400)
OTHER INFORMATION:
NAME/KEY: misc feature
LOCATION: (32163)..(32163)
OTHER INFORMATION: nucleotide 32163 is a single nucleotide polymorphism which can be
NAME/KEY: exon
LOCATION: (32401)..(32528)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (32529)..(33414)
OTHER INFORMATION:
NAME/KEY: misc feature
LOCATION: (32614)..(32614)
OTHER INFORMATION: nucleotide 32614 is a single nucleotide polymorphism which can be
NAME/KEY: exon
LOCATION: (33415)..(33597)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (33598)..(34314)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (34315)..(34598)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (34589)..(36404)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (36405)..(36523)
OTHER INFORMATION:
NAME/KEY: Intron
LOCATION: (36524)..(38341)
OTHER INFORMATION:
NAME/KEY: exon
LOCATION: (38342)..(38653)
OTHER INFORMATION:
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Genbank/AC004923
DATABASE ENTRY DATE: 1999-12-21
RELEVANT RESIDUES: (1)..(38653)
US-09-922-445-1
Query Match 77.1%; Score 16.2; DB 4; Length 38653;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GGGATTCTCTAGAGGAGGAG 21
|||||
DB 15818 GGGACTCGAGAGGAGGAG 15798
RESULT 5
US-09-313-294A-1436
Sequence 1436, Application US/09313294A
GENERAL INFORMATION:
APPLICANT: Laigudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14

NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 1436
LENGTH: 287
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700550945H1
NAME/KEY: unsure
LOCATION: 109-110
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-1436
Query Match 75.2%; Score 15.8; DB 4; Length 287;
Best Local Similarity 89.5%; Pred. No. 54;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 GATTCTCTAGAGGAGGAG 21
|||||
DB 215 GAATCTCTAGATGGAAGGAG 233
RESULT 6
US-09-641-638-13/c
Sequence 13, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-15
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 13
LENGTH: 955
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 478
OTHER INFORMATION: 10-33-175 : polymorphic base C or T
NAME/KEY: misc binding
LOCATION: 459..477
OTHER INFORMATION: 10-33-175.mis1
NAME/KEY: misc binding
LOCATION: 479..498
OTHER INFORMATION: 10-33-175.mis2, potential complement
NAME/KEY: primer bind
LOCATION: 304..322
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer bind
LOCATION: 705..723
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc binding
LOCATION: 466..490
OTHER INFORMATION: 10-33-175 potential probe
US-09-641-638-13
Query Match 75.2%; Score 15.8; DB 4; Length 955;

```
Best Local Similarity 89.5%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GGATTCCTAGGGAAGGA 20
Db 55 GGATTCCTAGGGAATGA 37

RESULT 7
US-09-641-638-14/c
; Sequence 14, Application US/09641638
; Patent No. 6432848
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLIETIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641.638
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 955
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 478
; OTHER INFORMATION: 10-33-211 : polymorphic base C or T
; NAME/KEY: misc_binding
; LOCATION: 458..477
; OTHER INFORMATION: 10-33-211.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 479..498
; OTHER INFORMATION: 10-33-211.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 268..286
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 669..687
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 466..490
; OTHER INFORMATION: 10-33-211 potential probe
; US-09-641-638-14

Query Match 75.2%; Score 15.8; DB 4; Length 955;
Best Local Similarity 89.5%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GGATTCCTAGGGAAGGA 20
Db 19 GGATTCCTAGGGAATGA 1

RESULT 8
US-09-292-542A-1/c
; Sequence 1, Application US/09292542A
; Patent No. 6531279
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
```

```
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Genomic Sequence Of The 5-Lipoxygenase-Activating Protein (FLAP)
; Patent No. 6531279
; FILE REFERENCE: GENSET 026A
; CURRENT FILING DATE: 1999-04-15
; CURRENT APPLICATION NUMBER: US/09/292.542A
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: US 60/081893
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: US 60/091314
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/123406
; PRIOR FILING DATE: 1999-03-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 43069
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..7708
; OTHER INFORMATION: potential 5' regulatory region
; NAME/KEY: misc_feature
; LOCATION: 36604..43069
; OTHER INFORMATION: potential 3' regulatory region
; NAME/KEY: exon
; LOCATION: 7709..7852
; OTHER INFORMATION: exon1
; NAME/KEY: exon
; LOCATION: 16236..16335
; OTHER INFORMATION: exon2
; NAME/KEY: exon
; LOCATION: 24227..24297
; OTHER INFORMATION: exon3
; NAME/KEY: exon
; LOCATION: 28133..28214
; OTHER INFORMATION: exon4
; NAME/KEY: exon
; LOCATION: 36128..36605
; OTHER INFORMATION: exon5
; NAME/KEY: misc_feature
; LOCATION: 7783..7795
; OTHER INFORMATION: ATG
; NAME/KEY: misc_feature
; LOCATION: 36288..36290
; OTHER INFORMATION: stop : TAA
; NAME/KEY: polyA_signal
; LOCATION: 36581..36586
; OTHER INFORMATION: AATAAA
; NAME/KEY: misc_feature
; LOCATION: 7008..8116
; OTHER INFORMATION: homology with sequence in ref genbank : M60470
; NAME/KEY: misc_feature
; LOCATION: 15995..16549
; OTHER INFORMATION: homology with sequence in ref genbank : M63259
; NAME/KEY: misc_feature
; LOCATION: 24059..24597
; OTHER INFORMATION: homology with sequence in ref genbank : M63260
; NAME/KEY: misc_feature
; LOCATION: 27873..28412
; OTHER INFORMATION: homology with sequence in ref genbank : M63261
; NAME/KEY: misc_feature
; LOCATION: 35977..36926
; OTHER INFORMATION: homology with sequence in ref genbank : M63262
; NAME/KEY: misc_feature
; LOCATION: 7613
; OTHER INFORMATION: diverging nucleotide deletion of a A in ref : M60470
; NAME/KEY: misc_feature
; LOCATION: 16347
; OTHER INFORMATION: diverging nucleotide G in ref : M63259
; NAME/KEY: misc_feature
; LOCATION: 16348
```

OTHER INFORMATION: diverging nucleotide A in ref : M63259
NAME/KEY: misc_feature
LOCATION: 24060
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63260
NAME/KEY: misc_feature
LOCATION: 24067
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63260
NAME/KEY: misc_feature
LOCATION: 27903
OTHER INFORMATION: diverging nucleotide deletion of a C in ref : M63261
NAME/KEY: misc_feature
LOCATION: 28327
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63261
NAME/KEY: misc_feature
LOCATION: 3851..4189
OTHER INFORMATION: 10-517
NAME/KEY: misc_feature
LOCATION: 4120..4390
OTHER INFORMATION: 10-518
NAME/KEY: misc_feature
LOCATION: 4373..4792
OTHER INFORMATION: 10-253
NAME/KEY: misc_feature
LOCATION: 4814..5043
OTHER INFORMATION: 10-499
NAME/KEY: misc_feature
LOCATION: 4956..5422
OTHER INFORMATION: 10-500
NAME/KEY: misc_feature
LOCATION: 5524..5996
OTHER INFORMATION: 10-522
NAME/KEY: misc_feature
LOCATION: 6218..6672
OTHER INFORMATION: 10-503
NAME/KEY: misc_feature
LOCATION: 6522..6790
OTHER INFORMATION: 10-504
NAME/KEY: misc_feature
LOCATION: 7120..7574
OTHER INFORMATION: 10-204
NAME/KEY: misc_feature
LOCATION: 7513..7933
OTHER INFORMATION: 10-32
NAME/KEY: misc_feature
LOCATION: 16114..16533
OTHER INFORMATION: 10-33
NAME/KEY: misc_feature
LOCATION: 24072..24425
OTHER INFORMATION: 10-34
NAME/KEY: misc_feature
LOCATION: 27978..28401
OTHER INFORMATION: 10-35
NAME/KEY: misc_feature
LOCATION: 36020..36465
OTHER INFORMATION: 10-36
NAME/KEY: misc_feature
LOCATION: 36318..36669
OTHER INFORMATION: 10-498
NAME/KEY: misc_feature
LOCATION: 38441..38840
OTHER INFORMATION: 12-629
NAME/KEY: misc_feature
LOCATION: 42233..42749
OTHER INFORMATION: 12-628 complement
NAME/KEY: allele
LOCATION: 3950
OTHER INFORMATION: 10-517-100 : polymorphic base S
NAME/KEY: allele
LOCATION: 4243
OTHER INFORMATION: 10-518-125 : polymorphic base K
NAME/KEY: allele
LOCATION: 4312
OTHER INFORMATION: 10-518-194 : polymorphic base R

NAME/KEY: allele
LOCATION: 4490
OTHER INFORMATION: 10-253-118 : polymorphic base R
NAME/KEY: allele
LOCATION: 4670
OTHER INFORMATION: 10-253-298 : polymorphic base S
NAME/KEY: allele
LOCATION: 4687
OTHER INFORMATION: 10-253-315 : polymorphic base Y
NAME/KEY: allele
LOCATION: 4968
OTHER INFORMATION: 10-499-155 : polymorphic base R
NAME/KEY: allele
LOCATION: 5140
OTHER INFORMATION: 10-500-185 : polymorphic base Y
NAME/KEY: allele
LOCATION: 5213
OTHER INFORMATION: 10-500-258 : polymorphic base K
NAME/KEY: allele
LOCATION: 5364
OTHER INFORMATION: 10-500-410 : polymorphic base R
NAME/KEY: allele
LOCATION: 5594
OTHER INFORMATION: 10-522-71 : polymorphic base R
NAME/KEY: allele
LOCATION: 6370
OTHER INFORMATION: 10-503-159 : polymorphic base K
NAME/KEY: allele
LOCATION: 6693
OTHER INFORMATION: 10-504-172 : polymorphic base W
NAME/KEY: allele
LOCATION: 6763
OTHER INFORMATION: 10-504-243 : polymorphic base M
NAME/KEY: allele
LOCATION: 7445
OTHER INFORMATION: 10-204-326 : polymorphic base R
NAME/KEY: allele
LOCATION: 7870
OTHER INFORMATION: 10-32-357 : polymorphic base M
NAME/KEY: allele
LOCATION: 16288
OTHER INFORMATION: 10-33-175 : polymorphic base Y
NAME/KEY: allele
LOCATION: 16347
OTHER INFORMATION: 10-33-234 : polymorphic base M
NAME/KEY: allele
LOCATION: 16383
OTHER INFORMATION: 10-33-270 : polymorphic base R
NAME/KEY: allele

Query Match 75.2%; Score 15.8; DB 4; Length 43069;
Best Local Similarity 89.5%; Pred. No. 1.12e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGATTCCTAGAGGAGGA 20
DB 15865 GGATTCCTTGGAGGAATGA 15847

RESULT 9
US-08-724-394A-16/c
Sequence 16, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereto

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/POCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3502 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..3502
OTHER INFORMATION: /note= "cDNA 32"
US-08-724-394A-16

Query Match 73.3%; Score 15.4; DB 2; Length 3502;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAA 17
DB 2924 GGGATTCCTAGAGAA 2908

RESULT 10
US-09-621-976-12965
; Sequence 12965, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 12965
; LENGTH: 288
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 286
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-12965

Query Match 72.4%; Score 15.2; DB 4; Length 288;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGGATTCCTAGAGGAAGGAG 21
DB 207 GGGATTCCTAGAGGAG 226
RESULT 11
US-09-621-976-10679/c
; Sequence 10679, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10679
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10679

Query Match 72.4%; Score 15.2; DB 4; Length 351;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAAGCA 20
DB 318 GGGATTCCTAGCGAAGCA 299

RESULT 12
US-09-621-976-2535/c
; Sequence 2535, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2535
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 232..402
US-09-621-976-2535

Query Match 72.4%; Score 15.2; DB 4; Length 463;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGATTCCTAGAGGAAGCA 20
DB 436 GGGATTCCTAGCGAAGCA 417

RESULT 13
US-09-621-976-18505/c
; Sequence 18505, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.

```
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18505
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-18505

Query Match      72.4%; Score 15.2; DB 4; Length 554;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GGATTCCTAGAGGGAAGGAG 21
      ||||| ||||| |||||
DB      543 GGATTCAGGAAGGAAGGAG 524

RESULT 14
US-09-252-991A-13569/c
; Sequence 13569, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13569
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-13569

Query Match      72.4%; Score 15.2; DB 4; Length 642;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GGGATTCTCTAGAGGGAAGGA 20
      ||||| ||||| |||||
DB      24 GGGATCCTTGAGTGAAGGA 5

RESULT 15
US-09-252-991A-12999/c
; Sequence 12999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12999
; LENGTH: 810
; TYPE: DNA
```

```
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-12999

Query Match      72.4%; Score 15.2; DB 4; Length 810;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GGGATTCTCTAGAGGGAAGGA 20
      ||||| ||||| |||||
DB      799 GGGTTTCCTTGAGGAAGGA 780

Search completed: March 25, 2004, 15:34:29
Job time : 6.13061 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 08:39:03 ; Search time 61.3218 Seconds
(without alignments)
6361.316 Million cell updates/sec

Title: US-09-963-285-1_COPY_223_231

Perfect score: 9
Sequence: 1 acaatgtt 9

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sv.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rtd.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	9	100.0	17	6	AR039289	AR039289 Sequence
C 2	9	100.0	17	6	AR039291	AR039291 Sequence
C 3	9	100.0	17	6	AR040261	AR040261 Sequence
C 4	9	100.0	17	6	AR040263	AR040263 Sequence
C 5	9	100.0	17	6	BD256634	BD256634 Regulatio
C 6	9	100.0	17	6	BD256635	BD256635 Regulatio
C 7	9	100.0	17	6	BD257082	BD257082 Regulatio
C 8	9	100.0	17	6	BD257083	BD257083 Regulatio
C 9	9	100.0	17	6	AR190012	AR190012 Sequence
C 10	9	100.0	17	6	AR190013	AR190013 Sequence
C 11	9	100.0	17	6	AR324989	AR324989 Sequence
C 12	9	100.0	17	6	AR324990	AR324990 Sequence
C 13	9	100.0	17	6	AX674725	AX674725 Sequence
C 14	9	100.0	17	6	AX737260	AX737260 Sequence
C 15	9	100.0	17	6	AX738366	AX738366 Sequence
C 16	9	100.0	17	6	AX759329	AX759329 Sequence
C 17	9	100.0	17	6	AX761772	AX761772 Sequence
C 18	9	100.0	18	6	BD261376	BD261376 Cadherin-
C 19	9	100.0	18	6	AX353253	AX353253 Sequence
C 20	9	100.0	18	6	AX353256	AX353256 Sequence
C 21	9	100.0	18	6	AX353258	AX353258 Sequence
C 22	9	100.0	18	6	AX353267	AX353267 Sequence
C 23	9	100.0	18	6	AX353272	AX353272 Sequence
C 24	9	100.0	18	6	AX363098	AX363098 Sequence
C 25	9	100.0	18	6	AX363101	AX363101 Sequence
C 26	9	100.0	18	6	AX363103	AX363103 Sequence
C 27	9	100.0	18	6	AX363112	AX363112 Sequence
C 28	9	100.0	18	6	AX363117	AX363117 Sequence
C 29	9	100.0	19	6	AX130438	AX130438 Sequence
C 30	9	100.0	19	6	AX130439	AX130439 Sequence
C 31	9	100.0	19	6	AX130440	AX130440 Sequence
C 32	9	100.0	19	6	AX130441	AX130441 Sequence
C 33	9	100.0	20	6	AR224678	AR224678 Sequence
C 34	9	100.0	20	6	AR224679	AR224679 Sequence
C 35	9	100.0	20	6	AX392100	AX392100 Sequence
C 36	9	100.0	20	6	AX355426	AX355426 Sequence
C 37	9	100.0	20	6	AX823621	AX823621 Sequence
C 38	9	100.0	22	6	AR024478	AR024478 Sequence
C 39	9	100.0	22	6	AR089048	AR089048 Sequence
C 40	9	100.0	22	6	AR089481	AR089481 Sequence
C 41	9	100.0	22	6	AR140684	AR140684 Sequence
C 42	9	100.0	22	6	AR428284	AR428284 Sequence
C 43	9	100.0	22	6	AX352505	AX352505 Sequence
C 44	9	100.0	22	6	BD023016	BD023016 Species-s
C 45	9	100.0	23	6	AX077239	AX077239 Sequence

ALIGNMENTS

RESULT 1
AR039289/c
LOCUS AR039289 17 bp DNA
DEFINITION Sequence 137 from patent US 5807743.
ACCESSION AR039289
VERSION AR039289.1 GI:5958652
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 17)
AUTHORS Stinchcomb,D.T. and McSwiggen,J.A.
TITLE Interleukin-2 receptor gamma-chain ribozymes
JOURNAL Patent: US 5807743-A 137 15-SEP-1998;
FEATURES Location/Qualifiers

linear PAT 29-SEP-1999

```

source 1..17
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAATGTT 9
Db 12 ACAAATGTT 4

RESULT 2
AR039291/c
LOCUS AR039291 17 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 139 from patent US 5807743.
ACCESSION AR039291
VERSION AR039291.1 GI:5958654
KEYWORDS
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Stinchcomb,D.T. and McSwiggen,J.A.
TITLE Interleukin-2 receptor gamma-chain ribozymes
JOURNAL Patent: US 5807743-A 139 15-SEP-1998;
FEATURES
Location/Qualifiers
1..17
source
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAATGTT 9
Db 11 ACAAATGTT 3

RESULT 3
AR040261/c
LOCUS AR040261 17 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1109 from patent US 5807743.
ACCESSION AR040261
VERSION AR040261.1 GI:5959624
KEYWORDS
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Stinchcomb,D.T. and McSwiggen,J.A.
TITLE Interleukin-2 receptor gamma-chain ribozymes
JOURNAL Patent: US 5807743-A 1109 15-SEP-1998;
FEATURES
Location/Qualifiers
1..17
source
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAATGTT 9
Db 12 ACAAATGTT 4

RESULT 4

```

```

AR040263/c
LOCUS AR040263 17 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1111 from patent US 5807743.
ACCESSION AR040263
VERSION AR040263.1 GI:5959626
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Stinchcomb,D.T. and McSwiggen,J.A.
TITLE Interleukin-2 receptor gamma-chain ribozymes
JOURNAL Patent: US 5807743-A 1111 15-SEP-1998;
FEATURES
Location/Qualifiers
1..17
source
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAATGTT 9
Db 11 ACAAATGTT 3

RESULT 5
BD256634/c
LOCUS BD256634 17 bp DNA linear PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION BD256634
VERSION BD256634.1 GI:33066404
KEYWORDS JP 2002541795-A/4427.
SOURCE unidentified
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 17)
AUTHORS Blatt,L., Zwick,M., Pavco,P. and McSwiggen,J.
TITLE Regulation of repressor genes using nucleic acid molecules
JOURNAL Patent: JP 2002541795-A 4427 10-DEC-2002;
COMMENT RIBOZYME PHARMACEUTICALS INC
OS Eukaryote
PN JP 2002541795-A/4427
PD 10-DEC-2002
PF 11-APR-2000 JP 2000611654
PR 12-APR-1999 US 60/129390
PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
C12N15/09,A61K38/00,A61K48/00,A61P43/00,A61P43/00,C12N5/10, PC
C12P21/02,
PC C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
C12R1:91),
PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
PC A61K37/02,
PC (C12N5/00,C12R1:91)
CC Regulation of repressor genes using nucleic acid molecules FH
Key source 1..17
Location/Qualifiers
FT source
/organism="Eukaryote".
FEATURES
Location/Qualifiers
1..17
source
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAATGTT 9

```

```

Db      12  ACAAATGTT 4
|||||
RESULT 6
BD256635/c
LOCUS      17 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION  BD256635
VERSION     BD256635.1  GI:33066405
KEYWORDS   JP 2002541795-A/4428.
SOURCE     unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE  1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE     Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4428 10-DEC-2002;
          RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
          PN JP 2002541795-A/4428
          PD 10-DEC-2002
          PF 11-APR-2000 JP 2000611654
          PR 12-APR-1999 US 60/129390
          PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
          C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
          C12P21/02,
          PC
          C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
          C12R1:91),
          PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
          PC A61K37/02,
          PC A61K37/02,C12R1:91)
          CC Regulation of repressor genes using nucleic acid molecules FH
          Key source 1..17
          Location/Qualifiers
          FT source 1..17
          /organism='Eukaryote'.
          FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      11  ACAAATGTT 3

RESULT 7
BD257082/c
LOCUS      17 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION  BD257082
VERSION     BD257082.1  GI:33066852
KEYWORDS   JP 2002541795-A/4875.
SOURCE     unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE  1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE     Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4875 10-DEC-2002;
          RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
          PN JP 2002541795-A/4875
          PD 10-DEC-2002
          PF 11-APR-2000 JP 2000611654
          PR 12-APR-1999 US 60/129390
          PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
          C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
          C12P21/02,
          PC
          C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
          C12R1:91),
          PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
          PC A61K37/02,
          PC A61K37/02,C12R1:91)
          CC Regulation of repressor genes using nucleic acid molecules FH
          Key source 1..17
          Location/Qualifiers
          FT source 1..17
          /organism='Eukaryote'.
          FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      11  ACAAATGTT 3

RESULT 8
BD257083/c
LOCUS      17 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION  BD257083
VERSION     BD257083.1  GI:33066853
KEYWORDS   JP 2002541795-A/4876.
SOURCE     unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE  1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE     Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4876 10-DEC-2002;
          RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
          PN JP 2002541795-A/4876
          PD 10-DEC-2002
          PF 11-APR-2000 JP 2000611654
          PR 12-APR-1999 US 60/129390
          PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
          C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
          C12P21/02,
          PC
          C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
          C12R1:91),
          PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
          PC A61K37/02,
          PC A61K37/02,C12R1:91)
          CC Regulation of repressor genes using nucleic acid molecules FH
          Key source 1..17
          Location/Qualifiers
          FT source 1..17
          /organism='Eukaryote'.
          FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      12  ACAAATGTT 4

RESULT 9
BD257083/c
LOCUS      17 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION  BD257083
VERSION     BD257083.1  GI:33066853
KEYWORDS   JP 2002541795-A/4876.
SOURCE     unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE  1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE     Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4876 10-DEC-2002;
          RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
          PN JP 2002541795-A/4876
          PD 10-DEC-2002
          PF 11-APR-2000 JP 2000611654
          PR 12-APR-1999 US 60/129390
          PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
          C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
          C12P21/02,
          PC
          C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
          C12R1:91),
          PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
          PC A61K37/02,
          PC A61K37/02,C12R1:91)
          CC Regulation of repressor genes using nucleic acid molecules FH
          Key source 1..17
          Location/Qualifiers
          FT source 1..17
          /organism='Eukaryote'.
          FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      12  ACAAATGTT 4

```

```

PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
C12N15/09,A61K38/00,A61K48/00,A61P43/00,A61P43/00,C12N5/10, PC
C12P21/02,
PC
C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
C12R1:91),
PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
PC A61K37/02,
PC A61K37/02,C12R1:91)
CC Regulation of repressor genes using nucleic acid molecules FH
Key source 1..17
Location/Qualifiers
FT source 1..17
/organism='Eukaryote'.
FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      12  ACAAATGTT 4

RESULT 8
BD257083/c
LOCUS      17 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION  BD257083
VERSION     BD257083.1  GI:33066853
KEYWORDS   JP 2002541795-A/4876.
SOURCE     unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE  1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE     Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4876 10-DEC-2002;
          RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
          PN JP 2002541795-A/4876
          PD 10-DEC-2002
          PF 11-APR-2000 JP 2000611654
          PR 12-APR-1999 US 60/129390
          PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
          C12N15/09,A61K38/00,A61K48/00,A61P43/00,A61P43/00,C12N5/10, PC
          C12P21/02,
          PC
          C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
          C12R1:91),
          PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
          PC A61K37/02,
          PC A61K37/02,C12R1:91)
          CC Regulation of repressor genes using nucleic acid molecules FH
          Key source 1..17
          Location/Qualifiers
          FT source 1..17
          /organism='Eukaryote'.
          FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      12  ACAAATGTT 4

RESULT 9
BD257083/c
LOCUS      17 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION  BD257083
VERSION     BD257083.1  GI:33066853
KEYWORDS   JP 2002541795-A/4876.
SOURCE     unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE  1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE     Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4876 10-DEC-2002;
          RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
          PN JP 2002541795-A/4876
          PD 10-DEC-2002
          PF 11-APR-2000 JP 2000611654
          PR 12-APR-1999 US 60/129390
          PI LAWRENCE BLATT, MICHAEL ZWICK, PAMELA PAVCO, JAMES MCSWIGGEN PC
          C12N15/09,A61K38/00,A61K48/00,A61P43/00,A61P43/00,C12N5/10, PC
          C12P21/02,
          PC
          C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
          C12R1:91),
          PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
          PC A61K37/02,
          PC A61K37/02,C12R1:91)
          CC Regulation of repressor genes using nucleic acid molecules FH
          Key source 1..17
          Location/Qualifiers
          FT source 1..17
          /organism='Eukaryote'.
          FT
FEATURES             source
            source
            1..17
            /organism='unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'
ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ACAAATGTT 9
      |||||||
      DB      12  ACAAATGTT 4

```



```
Qy 1 ACAATGTT 9
Db 11 ACAATGTT 3

RESULT 9
LOCUS AR190012/c 17 bp DNA PAT 20-APR-2002
DEFINITION Sequence 5500 from patent US 6346398.
ACCESSION AR190012
VERSION AR190012.1 GI:20235977
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Pavco,P., McSwiggen,J., Stinchcomb,D. and Escobedo,J.
TITLE Method and reagent for the treatment of diseases or conditions
related to levels of vascular endothelial growth factor receptor
JOURNAL Patent: US 6346398-A 5500 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..17
/mol_type="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAATGTT 9
Db 12 ACAATGTT 4

RESULT 10
LOCUS AR190013/c 17 bp DNA PAT 20-APR-2002
DEFINITION Sequence 5501 from patent US 6346398.
ACCESSION AR190013
VERSION AR190013.1 GI:20235978
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Pavco,P., McSwiggen,J., Stinchcomb,D. and Escobedo,J.
TITLE Method and reagent for the treatment of diseases or conditions
related to levels of vascular endothelial growth factor receptor
JOURNAL Patent: US 6346398-A 5501 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..17
/mol_type="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAATGTT 9
Db 12 ACAATGTT 4

RESULT 11
LOCUS AR324989/c 17 bp RNA PAT 17-AUG-2003
DEFINITION Sequence 2391 from patent US 6566127.
ACCESSION AR324989
VERSION AR324989.1 GI:33710797
KEYWORDS
SOURCE
ORGANISM Unknown.
```

```
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Pavco,P., McSwiggen,J.A., Stinchcomb,D.T. and Escobedo,J.
TITLE Method and reagent for the treatment of diseases or conditions
related to levels of vascular endothelial growth factor receptor
JOURNAL Patent: US 6566127-A 2391 20-MAY-2003;
FEATURES Location/Qualifiers
source 1..17
/mol_type="unknown"
/mol_type="unassigned RNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAATGTT 9
Db 12 ACAATGTT 4

RESULT 12
LOCUS AR324990/c 17 bp RNA PAT 17-AUG-2003
DEFINITION Sequence 2392 from patent US 6566127.
ACCESSION AR324990
VERSION AR324990.1 GI:33710798
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 17)
AUTHORS Pavco,P., McSwiggen,J.A., Stinchcomb,D.T. and Escobedo,J.
TITLE Method and reagent for the treatment of diseases or conditions
related to levels of vascular endothelial growth factor receptor
JOURNAL Patent: US 6566127-A 2392 20-MAY-2003;
FEATURES Location/Qualifiers
source 1..17
/mol_type="unknown"
/mol_type="unassigned RNA"

ORIGIN
Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAATGTT 9
Db 11 ACAATGTT 3

RESULT 13
LOCUS AX674725 17 bp DNA PAT 27-MAR-2003
DEFINITION Sequence 3170 from Patent WO03004526.
ACCESSION AX674725
VERSION AX674725.1 GI:293333073
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE 1
AUTHORS Telerman,A., Amson,R. and Tuijinder,M.
TITLE Sequences involved in phenomena of tumour suppression, tumour
reversion, apoptosis and/or resistance to viruses and their use as
medicines
JOURNAL Patent: WO 03004526-A 3170 16-JAN-2003;
FEATURES Molecular Engines Laboratories (FR)
source 1..17
/mol_type="unknown"
/mol_type="unassigned RNA"
```

/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAAATGTT 9
| | | | | | | | |
Db 9 ACAAAATGTT 17

RESULT 14

AX737260/c 17 bp DNA linear PAT 08-MAY-2003
LOCUS
DEFINITION Sequence 2850 from Patent WO03025177.
ACCESSION AX737260
VERSION AX737260.1 GI:30516548
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1
AUTHORS Telerman, A., Amson, R. and Tuijinder, M.
TITLE Sequences involved in phenomena of tumour suppression, tumour
reversion, apoptosis and/or resistance to viruses and the use
thereof as medicaments

JOURNAL Patent: WO 03025177-A 2850 27-MAR-2003;

FEATURES Molecular Engines Laboratories (FR)

source

1..17
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAAATGTT 9
| | | | | | | | |
Db 17 ACAAAATGTT 9

RESULT 15

AX738366 17 bp DNA linear PAT 08-MAY-2003
LOCUS
DEFINITION Sequence 3956 from Patent WO03025177.
ACCESSION AX738366
VERSION AX738366.1 GI:30517654
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1
AUTHORS Telerman, A., Amson, R. and Tuijinder, M.
TITLE Sequences involved in phenomena of tumour suppression, tumour
reversion, apoptosis and/or resistance to viruses and the use
thereof as medicaments

JOURNAL Patent: WO 03025177-A 3956 27-MAR-2003;

FEATURES Molecular Engines Laboratories (FR)

source

1..17
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 9; DB 6; Length 17;

Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAAATGTT 9
| | | | | | | | |
Db 9 ACAAAATGTT 17

Search completed: March 25, 2004, 12:49:45
Job time : 64.3218 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 08:39:03 ; Search time 1369.52 seconds
(without alignments)
6361.316 Million cell updates/sec

Title: US-09-963-285-1_COPY_223_423

Perfect score: 201

Sequence: 1 acaaatgtctccctgaagc.....ggattctctagagggaaggag 201

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rdt.*
- 36: em_htg_nam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	201	100.0	168656	9	AC009108	AC009108 Homo sapi
C 2	66	32.8	6021	10	MMHFHEAD1	Y08222 M.musculus
C 3	66	32.8	178416	10	AC127554	AC127554 Mus muscu
C 4	38	18.9	186021	9	AC136621	AC136621 Homo sapi
C 5	38	18.9	189923	9	AC026802	AC026802 Homo sapi
C 6	36.8	18.3	129727	9	AC084381	AC084381 Homo sapi
C 7	36.8	18.3	241394	2	AC131549	AC131549 Rattus no
C 8	36.8	18.3	261236	2	AC119381	AC119381 Rattus no
C 9	36.4	18.1	128680	2	AC023374	AC023374 Homo sapi
C 10	36.4	18.1	146376	2	AC009247	AC009247 Homo sapi
C 11	36.4	18.1	166236	9	AC022832	AC022832 Homo sapi
C 12	35.2	17.5	163579	2	AC104476	AC104476 Pan trogl
C 13	35.2	17.5	167671	5	BX000462	BX000462 Zebrafish
C 14	35.2	17.5	180971	9	AC079310	AC079310 Homo sapi
C 15	35.2	17.5	189932	2	BX649492	BX649492 Danio rer
C 16	35.2	17.5	195516	2	AC053540	AC053540 Homo sapi
C 17	35.2	17.5	273859	5	BX649333	BX649333 Zebrafish
C 18	35	17.4	137271	9	AC004552	AC004552 Homo sapi
C 19	34.8	17.3	174437	9	AC018470	AC018470 Homo sapi
C 20	34.6	17.2	96887	8	AC066689	AC066689 Arabidops
C 21	34.4	17.1	158151	2	AC103904	AC103904 Canis fam
C 22	34.4	17.1	161338	2	AC105372	AC105372 Canis fam
C 23	34.4	17.1	162407	2	AC073969	AC073969 Homo sapi
C 24	34.4	17.1	187543	9	AC068130	AC068130 Homo sapi
C 25	34	16.9	176345	4	AC104575	AC104575 Sus scrof
C 26	34	16.9	225962	2	AC094595	AC094595 Rattus no
C 27	34	16.9	231904	2	AC114208	AC114208 Rattus no
C 28	34	16.9	232945	2	AC118407	AC118407 Rattus no
C 29	34	16.9	237544	2	AC109858	AC109858 Rattus no
C 30	33.8	16.8	5062	5	BMU251223	AJ251223 Bungarus
C 31	33.8	16.8	151482	9	AC068765	AC068765 Homo sapi
C 32	33.8	16.8	163535	2	AC026876	AC026876 Homo sapi
C 33	33.8	16.8	189476	2	AC009893	AC009893 Homo sapi
C 34	33.8	16.8	339168	2	AC078987	AC078987 Homo sapi
C 35	33.6	16.7	193795	2	AC139167	AC139167 Rattus no
C 36	33.6	16.7	261260	2	AC110942	AC110942 Rattus no
C 37	33.6	16.7	314578	9	AC115349	AC115349 Rattus no
C 38	33.4	16.6	153133	9	AC091980	AC091980 Homo sapi
C 39	33.4	16.6	175154	9	AC093246	AC093246 Homo sapi
C 40	33.2	16.5	100310	2	FL7A13	AL929138 Mouse DNA
C 41	33.2	16.5	100984	10	AL929138	AL929138 Mouse DNA
C 42	33.2	16.5	162564	10	AC127575	AC127575 Mus muscu
C 43	33.2	16.5	181829	9	AC023425	AC023425 Homo sapi
C 44	33.2	16.5	184091	2	BX294393	BX294393 Mus muscu
C 45	33.2	16.5	197248	2	AC134907	AC134907 Mus muscu

ALIGNMENTS

RESULT 1
AC009108/c AC009108 168656 bp DNA linear PRI 29-OCT-2002
LOCUS Homo sapiens chromosome 16 clone RP11-46309, complete sequence.
DEFINITION AC009108
ACCESSION AC009108
VERSION AC009108.10 GI:24418066
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168656)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 168656)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 168656)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 29, 2002 this sequence version replaced gi:13786306.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
Location/Qualifiers
1. 168656
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RF11-46309"
ORIGIN
Query Match 100.0%; Score 201; DB 9; Length 168656;
Best Local Similarity 100.0%; Pred. No. 2.3e-54;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAAATGTTCTCCCTGAAGCCCTTCCCTGCCCAACACGACGACCACTTCCAAATTC 60
Db 78510 ACAAATGTTCTCCCTGAAGCCCTTCCCTGCCCAACACGACGACCACTTCCAAATTC 78451
Qy 61 TGCCCGGTGTAGCTTGTAAAGGGGTGTCTACTCTTCAGGAAAGTGGGAAAGG 120
Db 78450 TGCCCGGTGTAGCTTGTAAAGGGGTGTCTACTCTTCAGGAAAGTGGGAAAGG 78391
Qy 121 GATCTGATTATGAGGTGTGGAAGTAATAATATAGTCCACAAATAAACAACCTGTCC 180
Db 78390 GATCTGATTATGAGGTGTGGAAGTAATAATATAGTCCACAAATAAACAACCTGTCC 78331
Qy 181 GGGATCTAGAGGAGGAG 201
Db 78330 GGGATCTAGAGGAGGAG 78310
RESULT 2
MMMFHEAD1
LOCUS MMFH-1 6021 bp DNA linear ROD 14-MAY-1997
DEFINITION M.musculus MFH-1 gene.
ACCESSION Y08222
VERSION Y08222.1 GI:1869968
KEYWORDS mesenchyme fork head-1 protein; MFH-1 gene.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Miura, N., Iida, K., Kakinuma, H., Yang, X.-L. and Sugiyama, T.
TITLE Isolation of the mouse (MFH-1) and human (FKHL 14) mesenchyme fork head-1 genes reveals conservation of their gene and protein structures
JOURNAL Genomics 41 (3), 489-492 (1997)
MEDLINE 97312722
PUBMED 9169153
REFERENCE 2 (bases 1 to 6021)
AUTHORS Miura, N.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-1996) N. Miura, Akita University School of

Medicine, Department of Biochemistry, 1-1-1 Hondo, Akita 010, JAPAN
Location/Qualifiers
1. 6021
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129"
/db_xref="taxon:10090"
1485. 4345
2070. 3554
/gene="MFH-1"
2070. 3554
/gene="MFH-1"
/codon_start=1
/product="mesenchyme fork head-1 protein"
/protein_id="CAA69399.1"
/db_xref="GI:1869968"
/db_xref="GOA:Q61850"
/db_xref="SWISS-PROT:Q61850"
/translation="MQARVSDPNALGVVPLSEONYVRAAGSYGGMASPMGVYSGH
PEQYAGNRSYAPYHQPAKDLVKEPVSYIALITMAIONAPEKKTINLGIYQFIM
DRFFYRKNQGMQNSIRHNLNLNECEKVKVPRDDKKFGKSYMTLDPDSYMFENGSP
LRRRRFRKKDVPKDEERHLKEPSTTAKGAPGTVPADGPKERAKVYVKEAAS
PALEVITKVELTSPGALQASPSASSTPAGSPDGLPEHHAAPNGLPGFVSITMT
LRTSPGGDLSAAARAGLVVPLALPYAAPPAAYTPCAQGLLEAASGAYQCSMKA
MSLYTGAERPAHVCVPPALDEALSDHSPGSPGLGALNLAAGQEGALGASGHQHFG
HLHQAPPPAPQPPAPQAPATQATSWYLNHGGLSLPGLHTEATQQTFFNVREMPNS
HLGLDNSSLGESQVSNASQLPYRATPSLYRHAAPYSYDCTKY"
4320. 4325
polyA_signal
ORIGIN
Query Match 32.8%; Score 66; DB 10; Length 6021;
Best Local Similarity 73.1%; Pred. No. 2.8e-10;
Matches 155; Conservative 0; Mismatches 45; Indels 12; Gaps 5;
Qy 1 ACAAATGTTCTCCCTGAAGCCCTTCCCTGCCCAACACGACGACCACTTCCAAATTC 60
Db 195 ACAAATGTTCTCCCTGAAGCCCTTCCCTGCCCAACACGACGACCACTTCCAAATTC 254
Qy 61 TGCCCGGTGTAGCTTGTAAAGGGGTGTCTACTCTTCAGGAAAGTGGGAAAG 117
Db 255 TGTCGAGGCGATAGAGCAATCCGTAGGAGACATCTCGGTCTTCTAGGAAAGCGGACCGA 314
Qy 118 G--GGATCTGATTA-----TTGAGTGTGGAAGGATTAATAATCAGTCCACAAATAAAC 171
Db 315 GCAGGGATCCGATGACGACTGGAGATGTTGAAGGAATAAAT-ACCAGTCCCAATAAAC 373
Qy 172 AACTGT--CGGGATTCCTAGAGGAGGAG 201
Db 374 AACTGTCCCGGATTCCTAGAGGAGGAG 405
RESULT 3
AC127554 178416 bp DNA linear ROD 27-NOV-2003
LOCUS Mus musculus BAC clone RP24-323K23 from chromosome 8, complete
DEFINITION sequence.
ACCESSION AC127554
VERSION AC127554.4 GI:33457241
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 178416)
AUTHORS Cordes, M. and Haglund, K.
TITLE The sequence of Mus musculus BAC clone RP24-323K23
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 178416)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 178416)
AUTHORS McPherson, J.D. and Waterston, R.H.


```

/note="CpG_island (%GC=66.5, o/e=0.72, #CpGs=121)"
53367. .53484
/rpt_family="Alu"
54128. .54233
/rpt_family="Alu"
55215. .55240
/rpt_family="ERVK"
55241. .55364
/rpt_family="Alu"
55365. .55460
/rpt_family="ERVK"
56094. .56379
/rpt_family="B4"
59334. .59702
/rpt_family="B2"
60089. .60165
/rpt_family="ID"
60513. .60709
/rpt_family="B2"
60886. .61042
/rpt_family="B4"
61852. .61997
/rpt_family="MER1_type"
62247. .62549
/rpt_family="MaLR"
62716. .62764
/rpt_family="ID"
65656. .65793
/rpt_family="Alu"
65929. .66103
/rpt_family="MIR"
67019. .67319
/rpt_family="MER121"
67684. .67898

Query Match      32.8%; Score 66; DB 10; Length 178416;
Best Local Similarity 73.1%; Pred. No. 3.4e-10;
Matches 155; Conservative 0; Mismatches 45; Indels 12; Gaps 5;

Qy 1 ACAAATGTTCTCCCTGAAGCCCTCTCCCTGCCCAACAGCAGCAACTTCCAAATTC 60
Db 38322 ACAAATGTTTATCTGTCGCTTCTTCCCTACCCACCGGACCAACAACCTCCAGAAGT 38381

Qy 61 TGCCCGTG---TTTAGCTTGTAAAGGGTGTCTACTCTTCAGGAAAGTGGAAA 117
Db 38382 TCTCGAGGCATAGAGCCATTCGGTAGGGACATCTCGTGCTTCTGAGGAAGCGGACCA 38441

Qy 118 G--GGGATCTGATTA---TTGAGGTGTGGAAGATATAATAATCAGTCCACAAATAAAC 171
Db 38442 GCAGGATCCGATCAGCACTGGAGATGTTGAAGATAAAT-ACCAGTCCCAATAAAC 38500

Qy 172 AAATGT--CCGGATTCCTAGAGGGAAGAG 201
Db 38501 AAATGTCCCGGATTCCTAGAGGGAAGAG 38532

RESULT 4
AC136621/c AC136621 186021 bp DNA linear PRI 31-OCT-2003
LOCUS Homo sapiens chromosome 16 clone RP11-481F24, complete sequence.
DEFINITION AC136621
ACCESSION AC136621
VERSION AC136621.2 GI:38093724
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 186021)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186021)
AUTHORS

```

```

AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 186021)
AUTHORS Stanford Human Genome Center and Los Alamos National Laboratory.
CONSTRM DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 31, 2003 this sequence version replaced gi:24635938.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.sngc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
source
1. 186021
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-481F24"
ORIGIN
Query Match      18.9%; Score 38; DB 9; Length 186021;
Best Local Similarity 60.8%; Pred. No. 0.49;
Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy 72 AGCTTGTAAAGGGTGTCTCACTCTTCAGGAAAGTGGAAAGGGATCTGATAT 131
Db 2791 AGAATTGTTCAAGGAAGTCTCTCATGCTTAAGGAAATGATGAAAGTGAACATATAA 2732

Qy 132 TGAGTGTGGAAGGATATAATAATCAGTCCCAATAAACAA 173
Db 2731 AGAAGAGTGTAGAACACGTAATATGTAGATAAATAATAA 2690

```

```

RESULT 5
AC026802 188923 bp DNA linear PRI 25-OCT-2003
LOCUS Homo sapiens chromosome 16 clone CTD-2545G24, complete sequence.
DEFINITION AC026802
ACCESSION AC026802
VERSION AC026802.9 GI:37991843
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 188923)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188923)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 188923)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 188923)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 5 (bases 1 to 188923)
AUTHORS Stanford Human Genome Center and Los Alamos National Laboratory.

```

```
CONSRM      DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL     Submitted (25-OCT-2003) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 25, 2003 this sequence version replaced gi:27151362.
           Draft Sequence Produced by DOE Joint Genome Institute
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center and Los Alamos
           National Laboratory
           www.sngc.stanford.edu
           Quality: Phrap Quality >=40 100% of Sequence;
           Estimated Total Number of Errors is 0.
FEATURES    Location/Qualifiers
             1..188923
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="16"
             /clone="CTD-2545G24"
ORIGIN
Query Match      18.9%; Score 38; DB 9; Length 188923;
Best Local Similarity 60.8%; Pred.No. 0.49;
Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 72 AGCCTTGTTAAAGGGTGTCTCACTTCCTTCAGGAAAGTGGAAAGGGGATCTGATTAT 131
DB 145440 AGAATTGTTCAAGGAGTCTCATGCTTAAGGAAATGATGAAAGTGGAACTATATA 145499
QY 132 TGAGTGTGGAGGAATTAATTAATCAGTCACCAATTAACAA 173
DB 145500 AGACAGTGTGTAACACAGTAATATGTAGATAAATATAAA 145541

RESULT 6
AC084381/c
LOCUS       AC084381      129727 bp      DNA      linear      PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-20K20 from 7, complete sequence.
ACCESSION  AC084381
VERSION    AC084381.6 GI:15638716
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 129727)
           Sultana, J.E. and Waterston, R.
           Toward a complete human genome sequence
           Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE    99063792
PUBMED     9847074
REFERENCE  2 (bases 1 to 129727)
           Shah, N., Kozlowski, A. and Elliott, G.
           The sequence of Homo sapiens BAC clone RP11-20K20
           Unpublished (2001)
JOURNAL    3 (bases 1 to 129727)
           Waterston, R.H.
           Direct Submission
           Submitted (30-OCT-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  4 (bases 1 to 129727)
           Waterston, R.H.
           Direct Submission
           Submitted (18-SEP-2001) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  5 (bases 1 to 129727)
           Waterston, R.
           Direct Submission
           Submitted (09-JAN-2002) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT    On Sep 18, 2001 this sequence version replaced gi:15144331.
```

```
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0020K20
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRF/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACs1.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTB-67M9, 2000 bp overlap; the
clone sequenced to the right is RP11-115N4. Actual start of this
clone is at base position 29168 of CTB-67M9; actual end is at base
position 129727 of RP11-20K20.

Polymorphisms exist between RP11-20K20 and CTB-67M9. Data from
AC010764 was used to finish this clone, AC084381.

FEATURES
source
1..129727
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/maps="7"
/clone="RP11-20K20"
/clone_lib="RPC1-11"
1..88
/rpt_family="Li"
repeat_region 116..295
/rpt_family=" (TCCCC)n"
repeat_region 309..485
/rpt_family="Li"
repeat_region 1519..2146
/rpt_family="ERV1"
repeat_region 2530..2552
/rpt_family=" (TGGA)n"
repeat_region 3843..4010
/rpt_family="MIR"
repeat_region 4110..4398
/rpt_family="MaLR"
```

```

repeat_region 4467..4775
/rpt_family="Alu"
repeat_region 5480..5556
/rpt_family="(TA)n"
repeat_region 6384..6452
/rpt_family="MIR"
repeat_region 7703..7784
/rpt_family="(TA)n"
repeat_region 10148..10267
/rpt_family="MER1_type"
repeat_region 10261..10357
/rpt_family="MER1_type"
repeat_region 11105..11406
/rpt_family="Alu"
repeat_region 13246..13330
/rpt_family="AT-rich"
misc_feature 15570..16155
/notes="similar to Rattus norvegicus EST BF413169
(NID:g11401158)"
repeat_region 17320..17397
/rpt_family="L2"
repeat_region 18230..18657
/rpt_family="L1"
repeat_region 18658..18946
/rpt_family="Alu"
repeat_region 18947..19698
/rpt_family="L1"
repeat_region 19640..19682
/rpt_family="AT-rich"
repeat_region 19773..19800
/rpt_family="AT-rich"
misc_feature 21149..21594
/notes="match to EST BE925812 (NID:g10451888)"
misc_feature 21149..21424
/notes="similar to Mus musculus EST BF472109
(NID:g11541292)"
misc_feature 21149..21420
/notes="similar to Homo sapiens EST BF352373
(NID:g11311447)"
misc_feature 21149..21379
/notes="match to EST BE560547 (NID:g9804267)"
misc_feature 21193..21729
/notes="match to EST BE169347 (NID:g6632068)"
misc_feature 21361..21728
/notes="similar to Mus musculus EST BE850519
(NID:g10308858) uw19b05.y1"
misc_feature 21381..21800
/notes="match to EST AW501161 (NID:g7114430)"
misc_feature 21838..22775
/notes="match to EST AL535374 (NID:g12798867)"
misc_feature 22190..22851
/notes="match to EST BG260591 (NID:g12770407)"
misc_feature 22343..22770
/notes="match to EST A1204999 (NID:g3758051) an03b07.x1"
misc_feature 22580..22808
/notes="match to EST BE765363 (NID:g10195287)"
misc_feature 22814..23472
/notes="match to EST BG573517 (NID:g13581170)"
misc_feature 23007..23672
/notes="similar to Homo sapiens EST BE545373
(NID:g9774018)"
misc_feature 23049..23259
/notes="match to EST 244576 (NID:g573722)"
repeat_region 23381..23450
/rpt_family="CT-rich"
misc_feature 23491..24123
/notes="match to EST AL120642 (NID:g5926541)"
misc_feature 23523..23995
/notes="similar to Bos taurus EST AV663778 (NID:g9922808)"
misc_feature 23594..24395
/notes="match to EST AL565894 (NID:g12917721)"
misc_feature 23801..24550
/notes="similar to Rattus norvegicus EST BG671672

```

```

(NID:g13893771)"
misc_feature 23815..24249
/notes="match to EST BE007039 (NID:g8267272)"
misc_feature 23993..24269
/notes="similar to Mus musculus EST BB192528
(NID:g853147)"
misc_feature 24000..24577
/notes="match to EST BI253803 (NID:g14805587)"
misc_feature 24018..24197
/notes="match to EST AA465707 (NID:g2191229) aa32e03.s1"
misc_feature 24021..24275
/notes="similar to Mus musculus EST BE955751
(NID:g10599608)"
misc_feature 24025..24576
/notes="match to EST AA883846 (NID:g29933376) am21h10.s1"
misc_feature 24121..24572
/notes="similar to Mus musculus EST BG797711"

Query Match 18.3%; Score 36.8; DB 9; Length 129727;
Best Local Similarity 66.2%; Pred. No. 1.2;
Matches 53; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 21 CCTCTTCCTGCGCCACGACGACGACCTTCCAAATTCGCCGCTTTAGCCTTCTT 80
DB 93848 CTTCTTCCTGCGCCACGACGACCTTCCAAATTCGCCGCTTTAGCCTTCTT 80
QY 81 AAAGGGGTGCTCACTCCTT 100
DB 93788 AACGTGCTCTCTGAGTGATT 93769

RESULT 7
AC131549 241394 bp DNA linear HTG 13-NOV-2002
LOCUS Rattus norvegicus clone CH230-11F15, *** SEQUENCING IN PROGRESS
DEFINITION *** 2 unordered pieces.
ACCESSION AC131549
VERSION AC131549.3 GI:24941082
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 241394)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Dexamo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuwa,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawnsy,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

```


Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 241394)
Rat Genome Sequencing Consortium.
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241394)
Rat Genome Sequencing Consortium.
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:22855669.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GEBC
Center clone name: CH230-11F15
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 203660 bases at least Q40
Consensus quality: 209296 bases at least Q30
Consensus quality: 212200 bases at least Q20
Estimated insert size: 215943; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will

* be preserved.
* 1 239695: contig of 239695 bp in length
* 239696 239795: gap of unknown length
* 239796 241394: contig of 1599 bp in length.
FEATURES
Location/Qualifiers
source
1..241394
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-11F15"
misc_feature
1..2347
/notes="wgs contig"
misc_feature
31491..32938
/notes="wgs contig"
misc_feature
35824..36681
/notes="wgs contig"
misc_feature
36732..38223
/notes="wgs contig"
misc_feature
102229..106221
/notes="wgs contig"
misc_feature
131470..133327
/notes="wgs contig"
misc_feature
215653..217923
/notes="wgs contig"
ORIGIN
Query Match 18.3%; Score 36.8; DB 2; Length 241394;
Best Local Similarity 39.8%; Pred. No. 1-2;
Matches 74; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 3 AATGTTCTCCCTGAAGCCCTCTTCCTGCCAACAGACGACCACTTCCAAAATCTG 62
Db 49007 AGATGCTATCCCTTAAACCTCTTCTGCAACACCTCCCAAGCATCATAGTTGT 49066
QY 63 CCGTGTTTACCTGTGTAAGGGGTCTCTCACTCTTCAGGGAAAGTGGAAAAGGGA 122
Db 49067 AAGTACATTCCTTCATACCTGTATGACTTGTCTAGTTCTAGTTCTAATAAGTACA 49126
QY 123 TCTGATTATTGAGGTGTGGAAGGAATAAATAATCACTGCCACAAATAACAACTGCCGG 182
Db 49127 ACTGTTAAATGNN 49186
QY 183 GATTC 188
Db 49187 NNNNCC 49192
RESULT 8
AC119381 261236 bp DNA linear HTG 08-OCT-2002
LOCUS Rattus norvegicus clone CH230-161P20, *** SEQUENCING IN PROGRESS
DEFINITION ***
ACCESSION AC119381
VERSION AC119381.4 GI:228556286
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 261236)
AUTHORS
Muzny, D., Marie, M., Metzker, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschuler, S., Amin, A., Angiano, D., Anyanebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Dunn, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.J., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhwari, L., Louissegh, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, K., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwunna, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 261236)
Worley, K.C.

Direct Submission
Submitted (26-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 261236)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 14, 2002 this sequence version replaced gi:21903180.
The genome in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWHU
Center clone name: CH230-161P20
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 235487 bases at least Q40
Consensus quality: 238590 bases at least Q30
Consensus quality: 240652 bases at least Q20
Estimated insert size: 261654; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 261236: Contig of 261236 bp in length.
Location/Qualifiers
1..261236
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-161P20"
misc_feature
1..1435
/note="wgs_contig"
misc_feature
1486..3278
/note="wgs_contig"

FEATURES
source
Query Match 18.3%; Score 36.8; DB 2; Length 261236;
Best Local Similarity 54.4%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 3 AAATGTTCTCCCTGAAGCCCTCTTCCTCCGCCAACACAGCAGCACTCCAAAATTCGTG 62
DB 143400 AGATGCTATCCCTTAAACCCCTTTTCCTGAACACCCCTCCCAAGCATCCATAGTTGTT 143459
QY 63 CCCGTGTTTAGCTTGTAAAGGGGTGTCTCACTCCCTCAGGGAAGTGGGAAAGGGGA 122
DB 143460 AAGGTACATTGCCTTACACTGTATGCTGTCTAGTTCTAGTTAATAAGTACA 143519
QY 123 TCTGATTATTAGGTG 138
DB 143520 ACTGTTAAAAATGGTG 143535

RESULT 9
AC023374
LOCUS
DEFINITION
Homo sapiens clone RP11-22F4, *** SEQUENCING IN PROGRESS ***, 74
unordered pieces.
AC023374
VERSION
AC023374.2 GI:7657735
KEYWORDS
HTG; HTGS PHASE1
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 128680)
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens chromosome, clone RP11-22F4
JOURNAL
Unpublished
2 (bases 1 to 128680)
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavskiy, L., Bouckhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Landers, T., Larcocque, K., Lehoczy, J., Levine, R.,
 Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
 McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M.,
 Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.

TITLE

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 27, 2000 this sequence version replaced gi:6970507.

All repeats were identified using RepeatMasker:

Smit, A.F.P. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4113

Center clone name: 22_F_4

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1301: contig of 1301 bp in length
 1302 1401: gap of 100 bp
 1403 2463: contig of 1062 bp in length
 2464 2563: gap of 100 bp
 2564 3612: contig of 1049 bp in length
 3613 3712: gap of 100 bp
 3713 4785: contig of 1073 bp in length
 4786 4885: gap of 100 bp
 4886 5996: contig of 1111 bp in length
 5997 6096: gap of 100 bp
 6097 7466: contig of 1370 bp in length
 7467 7566: gap of 100 bp
 7567 8610: contig of 1044 bp in length
 8611 8710: gap of 100 bp
 8711 9913: contig of 1203 bp in length
 9914 10013: gap of 100 bp
 10014 11040: contig of 1027 bp in length
 11041 11140: gap of 100 bp
 11141 12424: contig of 1284 bp in length
 12425 13524: gap of 100 bp
 13525 13822: contig of 1298 bp in length
 13823 13922: gap of 100 bp
 13923 14934: contig of 1012 bp in length
 14935 15034: gap of 100 bp
 15035 16184: contig of 1150 bp in length
 16185 16284: gap of 100 bp
 16285 17291: contig of 1007 bp in length
 17292 17391: gap of 100 bp
 17392 18472: contig of 1081 bp in length
 18473 18572: gap of 100 bp
 18573 19721: contig of 1149 bp in length
 19722 19821: gap of 100 bp
 19822 21058: contig of 1237 bp in length
 21059 21158: gap of 100 bp
 21159 22325: contig of 1167 bp in length

22425: gap of 100 bp
 2243: contig of 118 bp in length
 2244: gap of 100 bp
 2245: contig of 1599 bp in length
 2424: gap of 100 bp
 24243: contig of 1611 bp in length
 25953: gap of 100 bp
 26053: gap of 100 bp
 27302: contig of 1249 bp in length
 27402: gap of 100 bp
 27403: contig of 1388 bp in length
 28990: gap of 100 bp
 30136: contig of 1246 bp in length
 30236: gap of 100 bp
 31616: contig of 1380 bp in length
 31716: gap of 100 bp
 3318: contig of 1802 bp in length
 3318: gap of 100 bp
 34937: contig of 1319 bp in length
 35037: gap of 100 bp
 36372: contig of 1335 bp in length
 36472: gap of 100 bp
 38358: contig of 1886 bp in length
 38458: gap of 100 bp
 39477: contig of 1019 bp in length
 39577: gap of 100 bp
 40629: contig of 1052 bp in length
 40729: gap of 100 bp
 41995: contig of 1266 bp in length
 42095: gap of 100 bp
 43379: contig of 1284 bp in length
 43479: gap of 100 bp
 44865: contig of 1386 bp in length
 44965: gap of 100 bp
 46815: contig of 1850 bp in length
 46915: gap of 100 bp
 48514: contig of 1539 bp in length
 48514: gap of 100 bp
 50670: contig of 2036 bp in length
 50770: gap of 100 bp
 51821: contig of 1051 bp in length
 51921: gap of 100 bp
 54151: contig of 2230 bp in length
 54251: gap of 100 bp
 5543: contig of 1392 bp in length
 5543: gap of 100 bp
 57332: contig of 1589 bp in length
 57432: gap of 100 bp
 58921: contig of 1489 bp in length
 59021: gap of 100 bp
 60563: contig of 1542 bp in length
 60563: gap of 100 bp
 62029: contig of 1366 bp in length
 62129: gap of 100 bp
 63463: contig of 1334 bp in length
 63563: gap of 100 bp
 65051: contig of 1488 bp in length
 65151: gap of 100 bp
 66808: contig of 1857 bp in length
 66908: gap of 100 bp
 68875: contig of 1967 bp in length
 68975: gap of 100 bp
 70346: contig of 1371 bp in length
 70446: gap of 100 bp
 72459: contig of 2013 bp in length
 72559: gap of 100 bp
 74054: contig of 1495 bp in length
 74154: gap of 100 bp
 75973: contig of 1819 bp in length
 76073: gap of 100 bp
 77770: contig of 1697 bp in length
 77870: gap of 100 bp
 79793: contig of 1923 bp in length
 79893: gap of 100 bp

* 79894	82188:	contig	of 2295 bp	in length	
* 82189	82288:	gap	of 100 bp	in length	
* 82289	84708:	contig	of 2420 bp	in length	
* 84709	84808:	gap	of 100 bp	in length	
* 84809	86868:	contig	of 2060 bp	in length	
* 86869	86968:	gap	of 100 bp	in length	
* 86969	89138:	contig	of 2167 bp	in length	
* 89139	89238:	gap	of 100 bp	in length	
* 89239	91440:	contig	of 2205 bp	in length	
* 91441	91541:	gap	of 100 bp	in length	
* 91542	93441:	contig	of 1901 bp	in length	
* 93442	93541:	gap	of 100 bp	in length	
* 93543	95044:	contig	of 1503 bp	in length	
* 95045	95144:	gap	of 100 bp	in length	
* 95146	96928:	contig	of 1781 bp	in length	
* 96929	97028:	gap	of 100 bp	in length	
* 97029	98950:	contig	of 1925 bp	in length	
* 98951	99050:	gap	of 100 bp	in length	
* 99051	101096:	contig	of 2046 bp	in length	
* 101097	101196:	gap	of 100 bp	in length	
* 101197	101901:	contig	of 705 bp	in length	
* 101902	102001:	gap	of 100 bp	in length	
* 102002	104578:	contig	of 2574 bp	in length	
* 104579	104678:	gap	of 100 bp	in length	
* 104679	107733:	contig	of 3058 bp	in length	
* 107734	107833:	gap	of 100 bp	in length	
* 107834	110783:	contig	of 2950 bp	in length	
* 110784	110883:	gap	of 100 bp	in length	
* 110884	112896:	contig	of 2013 bp	in length	
* 112897	112996:	gap	of 100 bp	in length	
* 112997	115653:	contig	of 2657 bp	in length	
* 115654	115753:	gap	of 100 bp	in length	
Query Match					18.1%; Score 36.4; DB 2; Length 128680;
Best local similarity					54.5%; Pred. No. 1.6;
Matches 73; Conservative					0; Mismatches 61; Indels 0; Gaps 0;
QY	65	CGTGTAGCTTGTAAAGGGTGTCTCACTCTTCAGGAAAGTGGGAAAGGGGATC	124		
Db	116191	CGTAGAGCTAGTGTATCTATCTGAGACTCACTGAATCAAAAGTATTTGGGAAAGGTTATC	116250		
QY	125	TGATTTATGAGTGTGGAGGAATTAATATCACTCCACAAATAACAACTGTCGGGA	184		
Db	116251	TGAAATGGATATATGGAGCAATATTTAATATTCCTCGTATTTATTTATGACTTAAA	116310		
QY	185	TTCTAGAGGGAAG	198		
Db	116311	TTGAGAGATTGATG	116324		
RESULT 10	AC009247/c				
LOCUS	1	146376 bp	DNA	linear	PRI 14-DEC-2001
DEFINITION	Homo sapiens 3q27 BAC RP11-410F19 (Roswell Park Cancer Institute				
ACCESSION	AC009247				
VERSION	AC009247.13	GI:17737024			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 146376)				
	Muzny, D.M., Adams, C., Bailey, M., Barberia, J., Blankenburg, K.,				
	Bodda, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,				
	Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,				
	David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,				
	Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,				
	Forcum-Tansey, J., Frantz, P., Ganesh, R., Garcia, D.K., Gorrell, J.H.,				
	Gorrell, L.L., Guevara, W., Harris, K., He, X., Hernandez, J.,				
	Hodgson, A., Hogue, M., Holloway, C., Kosak, H., Jackson, L.E.,				
	Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondrjewski, N., Kong, Y.,				
	Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O.,				

Lozano, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P.,
Mei, G., Moore, S., Moorish, I., Morgan, M., Morris, S., Nash, S.,
Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Osawa, G., Parish, B.,
Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D.,
Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H.,
Simen, M., Sparks, A., Stamps, A., Sucgang, R., Tabor, P., Taylor, T.,
Vasquez, L., Vinson, R., Vo, Q., Wabba, M., Watlington, S.,
Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,
Wrenford, G., Yu, W., Zhou, X., Naylor, S.L., Nelson, D. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 146376)
Worley, K.C.
Direct Submission
Submitted (08-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 146376)
Worley, K.C.
Direct Submission
Submitted (08-DEC-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 146376)
Worley, K.C.
Direct Submission
Submitted (30-MAR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 146376)
Worley, K.C.
Direct Submission
Submitted (14-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 14, 2001 this sequence version replaced gi:13489132.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

```

FEATURES
    Location/Qualifiers
        1..146376
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="RP11-410F19"
            789..825
                repeat_region
            /rpt_family="AT-rich"
            1371..1711
                misc_feature
            /note="Region: Similar to Hs#S1074034 0246g11.x1 Homo sapiens cDNA, AI079997"
            1737..1781
                repeat_region
            /rpt_family="CA)n"
            3500..3520
                repeat_region
            /rpt_family="AT-rich"
            join(4659..4721,6288..6513)
                misc_feature
            /note="Region: Similar to Hs#S1609703 DXFZp762L097_r1 Homo sapiens cDNA, AL121466"
            5272..5325
                repeat_region
            /rpt_family="GC-rich"
            6178..6278
                repeat_region
            /rpt_family="GC-rich"
            6495..6537
                repeat_region
            /rpt_family="C-rich"
            6695..6718
                repeat_region
            /rpt_family="GC-rich"
            7005..7142
                repeat_region
            /rpt_family="G-rich"
            8361..8487
                repeat_region
            /rpt_family="MIR"
            9046..9067
                repeat_region
            /rpt_family="(TTG)n"
            complement(9153..9513)
                repeat_region
            /rpt_family="THE1B"
            complement(10270..10517)
                repeat_region
            /rpt_family="AluJo"
            complement(10736..11062)
                repeat_region
            /rpt_family="MER1B"
            11101..11155
                repeat_region
            /rpt_family="MIR"
            11362..11452
                repeat_region
            /rpt_family="L1MB8"
            complement(11696..11997)
                repeat_region
            /rpt_family="L2"
            12202..12827
                repeat_region
            /rpt_family="HAL1"
            12980..13267
                repeat_region
            /rpt_family="AluJo"
            13473..13671
                repeat_region
            /rpt_family="L2"
            complement(14040..14312)
                repeat_region
            /rpt_family="Aluub"
            complement(14495..14807)
                repeat_region
            /rpt_family="AluSp"
            complement(14897..15005)
                repeat_region
            /rpt_family="MER5A"
            complement(15090..15172)
                repeat_region
            /rpt_family="L1MD3"
            complement(15162..15221)
                repeat_region
            /rpt_family="L1MC4"
            15222..15533
                repeat_region
            /rpt_family="AluSc"
            complement(15534..15650)
                repeat_region
            /rpt_family="L1MC4"
            15651..15954
                repeat_region
            /rpt_family="AluYa5"
            complement(15955..16485)
                repeat_region
            /rpt_family="L1MC4"
            complement(16586..16678)
                repeat_region
            /rpt_family="AluSc/x"
            complement(16679..16966)
                repeat_region
            /rpt_family="L1MC4"

repeat_region 17043..17351
rpt_family="AluSg1"
misc_feature 17320..17410
/function="Low Coverage"
repeat_region 17376..17407
/rpt_family="(TTTG)n"
complement(17408..17699)
rpt_family="AluSx"
complement(17700..17843)
/rpt_family="L1MC4"
complement(17878..18184)
/rpt_family="AluSg"
complement(18422..18733)
/rpt_family="AluY"
complement(18919..20097)
/rpt_family="L1ME"
20107..20275
/rpt_family="L1PAL3"
20276..20580
/rpt_family="AluSx"
20581..20741
/rpt_family="L1PAL3"
20758..20865
/rpt_family="HAL1"
20879..20913
/rpt_family="(TTTA)n"
complement(20914..21210)
/rpt_family="AluSg"
21311
complement(21253..21311)
/rpt_family="L1R33"
complement(21334..21422)
/rpt_family="MER2"
21423..21723
/rpt_family="MER2"

Query Match 18.1% Score 36.4; DB 9; Length 146376;
Best Local Similarity 66.7%; Pred. No. 1.6;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 7 GTTCTCCCTGAGCCCTCTCCCTGCCCAACAGACCACTCCAAAATTCGCCG 66
    |||||
Db 61915 GTTCTCCCTGAGCCCGTTTCTCTCTCCCTCTCTGAAACATGCTAAATTTGACTG 61856

Qy 67 TGTTTAGCCTTGTTAAAG 84
    |||||
Db 61855 TGCTTCATTGTCATAG 61838

RESULT 11
AC022832/c
LOCUS AC022832 166236 bp DNA linear PRI 11-DEC-2001
DEFINITION Homo sapiens chromosome 8, clone RP11-145015, complete sequence.
ACCESSION AC022832
VERSION AC022832.13 GI:17488715
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166236)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-145015
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166236)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,B., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

```

McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 166236)

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farrow, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Headford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Traversi, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, M., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (02-SEP-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 166236)

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farrow, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Headford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Traversi, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, M., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 11, 2001 this sequence version replaced gi:15421017.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6140

Center clone name: 145_O_15

FEATURES
source
1. 166236
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-145015"
/clone_lib="RP11-11 Human Male BAC"

1. 625

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```

repeat_region complement(22793..23107)
repeat_region /rpt family="LTR33"
repeat_region complement(23947..24097)
repeat_region /rpt family="L1MEC"
repeat_region 24392..24463
repeat_region /rpt family="MERSA"
repeat_region complement(24860..25243)
repeat_region /rpt family="THEIC"
repeat_region complement(26003..26306)
repeat_region /rpt family="AluUb"
repeat_region 27314..27455
repeat_region /rpt family="L1ME1"
repeat_region 27456..27742
repeat_region /rpt family="AluSq"
repeat_region 27745..27780
repeat_region /rpt family="TAAA)n"
repeat_region 27781..27836
repeat_region /rpt family="L1ME1"
repeat_region complement(27837..28860)
repeat_region /rpt family="Tigger3b"
repeat_region 28861..29160
repeat_region /rpt family="Aluy"
repeat_region complement(29161..29350)
repeat_region /rpt family="Tigger3b"
repeat_region 29351..29557
repeat_region /rpt family="L1ME1"
repeat_region 29458..29604
repeat_region /rpt family="L1MAL0"
repeat_region complement(30649..30674)
repeat_region /rpt family="MIR"

Query Match 18.1%; Score 36.4; DB 9; Length 166236;
Best Local Similarity 54.5%; Pred. No. 1.6;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 65 CGTGTTTACCTGTTAAAGGGGTCTCCTCCTTCAGGAAGTGGAAAAGGGATC 124
DB 57908 CGTAGAGAGCTAGGTCAATCTGAGACTCACTGAATCAAGTATTTCGGAAGGTTATC 57849
QY 125 TGATTATTGAGTGGGAAGGAATAATAATACAGTCCCAATAAACAACCTGTCGGGA 184
DB 57848 TGAATATGATATATGGAGCAATATTATATTCCTCGTAAATATTATGACCTTAAA 57789
QY 185 TTCCTAGAGGGAAG 198
DB 57788 TTCGAAGATTGATG 57775

RESULT 12
AC104476/c
LOCUS Pan troglodytes clone RP43-139N13, WORKING DRAFT SEQUENCE, 3
DEFINITION ordered pieces.
ACCESSION AC104476
VERSION AC104476.2 GI:23266279
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE Pan troglodytes (chimpanzee).
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 163579)
AUTHORS Akter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Carliaga,K., Coleman,B., Dietrich,N.L., Granite,S., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R.,
Karline,E., Lari,C., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L.,
Maduro,V.B., Margulies,E.H., Mastello,C., Maskeri,B.,
Masrion,S.D., McCloskey,J.C., McDowell,J., Paquerigan,C.,
Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N.,
Schueler,M.G., Sison,C., Stantropop,S., Thomas,J.W., Thomas,P.J.,
Touchman,J.W., Vogt,J.L., Walker,M., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163579)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
REFERENCE 3 (bases 1 to 163579)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2002) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
COMMENT On Sep 21, 2002 this sequence version replaced gi:17530718.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: cez
Center clone name: 139N13

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 163300 bases at least Q40
 Consensus quality: 163363 bases at least Q30
 Consensus quality: 163375 bases at least Q20
 Insert size: 117000; agarose-fp
 Insert coverage: 163379; sum-of-contigs
 Quality coverage: 12.68x in Q20 bases; agarose-fp
 Quality coverage: 9.08x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 60466: contig of 60466 bp in length
 60467 60566: gap of unknown length
 60567 69329: contig of 8763 bp in length
 69330 69429: gap of unknown length
 69430 163579: contig of 94150 bp in length.

FEATURES

source
 1..163579
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="RP43-139N13"
 /clone_lib="RP43"
 1..60466
 /notes="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 1..30537
 /notes="clone overlaps with GenBank Accession Number AC104154 clone RP43-108N2 (center project name cez)"

TITLE

```

misc_feature      60567..69329
                  /note="assembly_fragment"
misc_feature      69430..163579
                  /note="assembly_fragment"
                  clone_end:T7
                  vector_side:right"

ORIGIN
Query Match      17.5%; Score 35.2; DB 2; Length 163579;
Best Local Similarity 65.0%; Pred.No.4;
Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 21 CCTCTCCCTGCCCAACAGACACCACTTCCAAAATTCGCCGGTGTAGCCTTGT 80
Db 163237 CCTCTCCCTGCCCAACAAATCAGCACCTCTGAATAGGCGCAGAAATCCTCATGTT 163178

QY 81 AAAGGGGTCTCACTCCTT 100
Db 163177 AACATGCTCTCTGAGTGATT 163158

RESULT 13
BX000462/c
LOCUS
DEFINITION      Zebrafish DNA sequence from clone DKEY-114F6 in linkage group 1,
                  complete sequence.
ACCESSION      BX000462
VERSION        BX000462.10 GI:34368570
KEYWORDS      HTG.
SOURCE        Danio rerio (zebrafish)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                  Cypriniformes; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 167671)
AUTHORS        Howden, P.
TITLE          Direct Submission
JOURNAL        Submitted (30-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
                  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                  zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                  On Aug 30, 2003 this sequence version replaced gi:33414474.
COMMENT        ----- Genome Center
                  Center: Wellcome Trust Sanger Institute
                  Center code: SC
                  Web site: http://www.sanger.ac.uk
                  Contact: zfish-help@sanger.ac.uk
                  -----
                  During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
                  together with a note of the overlapping clone name. Note that the
                  variation annotation may not be found in the sequence submission
                  corresponding to the overlapping clone, as we submit sequences with
                  only a small overlap as described above.
                  This sequence was finished as follows unless otherwise noted: all
                  regions were either double-stranded or sequenced with an alternate
                  chemistry or covered by high quality data (i.e., phred quality >=
                  30); an attempt was made to resolve all sequencing problems, such
                  as compressions and repeats; all regions were covered by at least
                  one plasmid subclone or more than one M13 subclone; and the
                  assembly was confirmed by restriction digest, except on the rare
                  occasion of the clone being a YAC.
                  The following abbreviations are used to associate primary accession
                  numbers given in the feature table with their source databases:
                  Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
                  on the WORMPEP database can be found at
                  http://www.sanger.ac.uk/Projects/Celegans/wormpep/
                  Zebrafish pUC subclones occasionally display inconsistency over the
                  length of mononucleotide A/T runs and conserved TA repeats. Where
                  this is found the longest good quality representation will be
                  submitted.
                  Repeat names beginning 'Dr' were identified by the Recon repeat
                  discovery system (Zhirong Bao and Sean Eddy, submitted), and those
                  beginning 'drr' were identified by Rick Waterman (Stephen Johnson

```

```

lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-114F6
is from a Zebrafish BAC library
VECTOR: pindigoBAC-S.
Location/Qualifiers
1..167671
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-114F6"
/clone_lib="DarioKey"

FEATURES
source
Query Match      17.5%; Score 35.2; DB 5; Length 167671;
Best Local Similarity 53.7%; Pred.No.4;
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 56 AATTCGCGCGTGTAGCCTTGTAAAGGGGTGTCTCCTCTCAGGGAAGTGGAA 115
Db 15112 AATTATGCTGCTGATTAGCTTAATTAAGCTTAGCTTATCCACTTAGGATTAGGGGCTA 15053

QY 116 AAGGGGATCTGATTATTGAGGTGTGAGGAGTAATAATCAGTCCACAAATATAACAAAC 175
Db 15052 AAGTAAAGCTGATTAATATGAGTGTGTTTCAGAAATCATTAAGTAAATCCCTGGTC 14993

QY 176 TGCCCGGATTCCTAG 191
Db 14992 TGCCCATGAGACAGAG 14977

RESULT 14
AC079310
LOCUS
DEFINITION      Homo sapiens 12 BAC RP11-850F7 (Roswell Park Cancer Institute Human
                  BAC Library) complete sequence.
ACCESSION      AC079310
VERSION        AC079310.20 GI:12000443
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 180971)
AUTHORS        Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
                  Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbara, J.,
                  Benton, J., Bivaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
                  Bowie, S., Brive, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
                  Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
                  Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
                  Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D.,
                  Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
                  Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
                  Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
                  Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
                  Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,
                  Flag, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,
                  Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W.,
                  Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,
                  Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O.,
                  Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homsi, F.,
                  Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E.,
                  Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
                  Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
                  Kravtsov, J., Kurehi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C.,
                  Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
                  Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
                  Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R.,
                  Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P.,
                  Meador, M., Mel, G., Merscher, K., Metzger, M., Miller, A., Miner, G.,
                  Minor, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M.,
                  Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,
                  Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
                  Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,

```


Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoshkari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Kucheriapathi, R. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 180971)

Worley, K.C.

Direct Submission

Submitted (27-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 180971)

Worley, K.C.

Direct Submission

Submitted (01-JAN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 180971)

Worley, K.C.

Direct Submission

Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 180971)

Worley, K.C.

Direct Submission

Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jan 1, 2001 this sequence version replaced gi:11968205.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----

Contig length: 180971

Phrap values in estimate: 180205

Average error rate (BCM-Phrap estimate): 8.90355e-06

Fraction of Phrap values less than 40 : 0.0075858

Number of consensus changing edits: 50

Number of N's in consensus : 0

----- Consensus changing edits -----

Position	Original+Context	Edited+Context
445	aaagtgggg(n)aaagtggcag	aaagtgggg(g)aaagtggcag
5307	tgtctcaaaa(n)aaataataa	tgtctcaaaa(t)aaataataa
5317	tgaagatat(n)catatata	tgaagatat(t)catatata
5855	atataatat(n)ttttttgt	atataatat(a)ttttttgt
5881	acagagtta(n)ttttttgt	acagagtta(c)ttttttgt
5882	cagagttta(n)ttttttgt	cagagttta(c)ttttttgt
5953	gggtccagc(n)attcttcgc	gggtccagc(a)attcttcgc
5959	agcnaattct(n)ctgcctcag	agcnaattct(t)ctgcctcag
5974	ctcagctcc(n)agatgggtg	ctcagctcc(t)agatgggtg
6033	attttagta(n)agatgggtg	attttagta(g)agatgggtg
6084	cctgacctca(n)gfgntctgc	cctgacctca(g)gfgntctgc
6088	acctcangtg(n)tttgcnnn	acctcangtg(a)tttgcnnn
6095	ggnctctgcc(n)nnnnnctc	ggnctctgcc(t)nnnnnctc
6096	gntctgcnn(n)nnnnnctc	gntctgcnn(g)nnnnnctc
6097	ntctgcnnn(n)nnnnctcc	ntctgcnnn(c)nnnnctcc
6098	ttctgcnnn(n)nnnnctcc	ttctgcnnn(t)nnnnctcc
6100	ctgcnnnnn(n)nnnnctcc	ctgcnnnnn(t)nnnnctcc
6101	tgccnnnnn(n)nnnnctcc	tgccnnnnn(g)nnnnctcc
6102	gcnnnnnn(n)nnnnctcc	gcnnnnnn(g)nnnnctcc
6735	gtgcagtgga(n)ccagatcgc	gtgcagtgga(g)ccagatcgc
22211	tgcaaaaggc(n)tgatcctgc	tgcaaaaggc(c)tgatcctgc
31019	acgggttttt(n)ccatattggc	acgggttttt(c)ccatattggc
31033	tattggccag(n)aaagtctcgn	tattggccag(g)aaagtctcgn
31043	naagctctcg(n)agatgcatt	naagctctcg(g)agatgcatt
31083	agtcacagaag(n)cccatctct	agtcacagaag(a)cccatctct
31124	cagacccttt(n)tatcccaaac	cagacccttt(c)tatcccaaac
31239	taaaagagat(n)cgctgtctcc	taaaagagat(a)cgctgtctcc
32343	gtcagcagtg(n)catagacctt	gtcagcagtg(a)catagacctt
43188	agaaaaaaa(n)ccactctctc	agaaaaaaa(a)ccactctctc
54549	atagggaata(n)aatcaatctg	atagggaata(a)aatcaatctg
62663	gtcaaaagtg(n)ttctctgca	gtcaaaagtg(g)ttctctgca
62664	tcaaaagtgn(n)ttctctgca	tcaaaagtg(t)ttctctgca
62669	agcgmntct(n)tgmnccat	agcgmntct(c)tgmnccat
62672	gnttctctg(n)cccatctcac	gnttctctg(c)cccatctcac
6273	nnctctctg(n)cccatctcac	nnctctctg(a)cccatctcac
84783	aggtttttga(n)ttgaaattgt	aggtttttga(a)ttgaaattgt
84787	ttttgaaatt(n)aatgttgtt	ttttgaaatt(a)aatgttgtt
107633	aactaaataa(n)ttatgcaatc	aactaaataa(a)ttatgcaatc
107634	actaaataa(n)ttatgcaatc	actaaataa(a)ttatgcaatc
118957	tgttggccag(n)ctggctctga	tgttggccag(g)ctggctctga
119521	agctctctat(n)tgatggagga	agctctctat(c)tgatggagga
137842	taattctctg(n)tcaggtctca	taattctctg(t)tcaggtctca
156893	ctaaataaat(n)ttactctgaa	ctaaataaat(c)ttactctgaa
163882	taattctctg(n)ctgtgtgaa	taattctctg(c)ctgtgtgaa
163892	acctgtgtga(n)aatctacttc	acctgtgtga(a)aatctacttc
172348	taaaataaac(n)ggccttttaa	taaaataaac(c)ggccttttaa
172474	ggagcgaat(n)acaaatgggc	ggagcgaat(t)acaaatgggc
172542	tgtaaaaaac(n)cacctatata	tgtaaaaaac(c)cacctatata
176104	caaaagtcag(n)gratcatctt	caaaagtcag(t)gratcatctt

----- Distribution of Quality < 40 Bases -----

1000
900
800

700
bases 600
500
400
300
200
100
0

5 10 15 20 25 30 35 40
Phrap Value Range

Query Match 17.5%; Score 35.2; DB 9; Length 180971;
Best Local Similarity 52.8%; Pred. No. 4.1;
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 21 CCTTTCCTGCGCCACACAGACAGCAACTCCAAAATTCTGCCCGTTTACCGTTGTT 80
DB 22236 CCTATCCCTTGGCTTACCTGGGATGACTGCTCACTGTCAGCACAGGGTGCAGGGT 22295

QY 81 AAGGGGTGCTCACTCCTTCAGGGAAGTGGGAAGGGGATCTGATTATTGAGGTG 140
DB 22296 GGGCGCGCGCTCCGATCTGAGGGAAGGTGGGAATGGGAAGAGTCAAGTACAGTG 22355

QY 141 GAAGGAATAAATATCATGTCACCA 164
DB 22356 GAAGCATCAAGCCCCACCACAAA 22379

RESULT 15
BX649492
LOCUS
DEFINITION Danio rerio clone DKEY-13P3, WORKING DRAFT SEQUENCE, 2 unordered pieces.
ACCESSION BX649492
VERSION BX649492.4 GI:38091215
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 189932)
Bates,K.
REFERENCE Direct Submission
AUTHORS Submitted (05-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
JOURNAL On Oct 30, 2003 this sequence version replaced gi:37988067.
COMMENT -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

Project Information
Center project name: zkl3p3

Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 189785 bases at least Q40
Consensus quality: 189794 bases at least Q30
Consensus quality: 189816 bases at least Q20
Insert size: 189832; sum-of-contigs
Insert size: 193531; 3.6% error; agarose-fp
Quality coverage: 19.57x in Q20 bases; sum-of-contigs Quality coverage: 20.58x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 112060: contig of 112060 bp in length
* 112061 112160: gap of 100 bp
* 112161 189932: contig of 77772 bp in length.

FEATURES
Location/Qualifiers
Source 1..189932
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-13P3"
/clone_lib="DanioKey"
misc_feature 1..112060
/note="assembly_fragment:03464.0"
misc_feature 112161..189932
/note="assembly_fragment:05810"

ORIGIN
Query Match 17.5%; Score 35.2; DB 2; Length 189932;
Best Local Similarity 53.7%; Pred. No. 4.1;
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 56 AATTCGCGCGTGTGTTAGCCTTGTAAAGGGGTGTCCTCCTTCAGGGAAGTGGAA 115
DB 64914 AATTATGCTGCTGATTAGCCCTTAATTAAAGCTTAGCTTATCAGTATAGGATAGGGGCTA 64973

QY 116 AAGGGGATCTGATTATTGAGGTGTGAAGGAATAAATAATCAGTCCACAAATAACAAC 175
DB 64974 AAGTAAGCCTGATTATATATGAGTGTGTTTCAGAAATCACTACTACTATCCCTGCTC 65033

QY 176 TGTCGGGATTCCTAG 191
DB 65034 TGCCCATGAGACAGAG 65049

Search completed: March 25, 2004, 12:50:05
Job time : 1375.52 secs

Result No.	Score		Query		Length	DB	ID	Description
	Match	%	Match	%				
1	201	100.0	6458	6	ABX48984	6	ABX48984	Genomic D
2	66	32.8	6021	6	ABX48986	6	ABX48986	Genomic D
3	32.4	16.1	5862	7	ACC90899	7	ACC90899	Human GGD
4	31.6	15.7	5983	4	ABX84857	4	ABX84857	Human imm
5	31.4	15.6	296	7	ABX81870	7	ABX81870	Corn ear-
6	31.4	15.6	22700	4	AAK70746	4	AAK70746	Human imm
7	31.2	15.5	213456	7	AAI54072	7	AAI54072	Genomic D
8	31	15.4	1830	6	AB211734	6	AB211734	Human pol
9	31	15.4	110000	7	ABD53223	3	Continuation (4 of	
10	30.8	15.3	360	7	ABX41193	7	ABX41193	Bovine ES
11	30.4	15.1	451	4	ABX50387	4	ABX50387	Human liv
12	30.4	15.1	3574	2	AA556241	2	AA556241	Mouse JMY
13	30.2	15.0	1466	3	AAA72296	3	AAA72296	Drosophil
14	30.2	15.0	2075	3	AA949412	3	AA949412	Drosophil
15	30.2	15.0	3548	8	AB222180	8	AB222180	Drosophil
16	30.2	15.0	9303	7	AB567485	7	AB567485	Human sec
17	30.2	15.0	9303	7	AB273903	7	AB273903	Secreted
18	30	14.9	4466	4	ABX69983	4	ABX69983	Human imm
19	29.8	14.8	274	4	ABX86185	4	ABX86185	Corn ear-
20	29.6	14.7	516	6	ABN68994	6	ABN68994	Streptoco
21	29.6	14.7	110000	6	ABN71527	12	Continuation (13 o	
22	29.6	14.7	142299	9	ADD50651	9	ADD50651	BAC seque
23	29.4	14.6	142299	9	AA221329	9	AA221329	Human sec

XX Sequence 6021 BP; 1350 A; 1675 C; 1642 G; 1354 T; 0 U; 0 Other;
Query Match 32.8%; Score 66; DB 6; Length 6021;
Best Local Similarity 73.1%; Pred. No. 1e-11;
Matches 155; Conservative 0; Mismatches 45; Indels 12; Gaps 5;

QY 1 ACNATGTTCCCTGAGCCCTCTTCCCTCCCAACGACGACCACTCCAAATTC 60
DB 195 ACAATGTTTATCTGTCGCTTCTCCCTACCAACCGGACCAACTCCAGAGGT 254
QY 61 TGCCCGTG---TTTAGCCTTGTAAAGGGGTGTCTCACTCTTCAGGGAAGTGGGAAA 117
DB 255 TCTGCGAGGATAGAGCCATTCCGTAGGAGACATCTCGTGCTTCTGAGGAGCGGACCGA 314
QY 118 G--GGATCTGATTA----TTGAGGTGTGGAGGAATAAATCACTCCACAAATAAC 171
DB 315 GCAGGGATCCGATGACGACTGGAGATGTTGAAGGAATAAAT-ACCAGTCCACAAATAAC 373
QY 172 AAACCTGT--CGGGATTCCTAGAGGGAAGGAG 201
DB 374 AAACCTGTCCCGGATTCCTAGAGGGAAGGAG 405

RESULT 3
ACC90589
ID ACC90589 standard; DNA; 6962 BP.
AC ACC90589;
XX
DT 12-AUG-2003 (first entry)
XX
DE Human CGDD-12 encoding DNA.
KW Human; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
KW Huntington's disease; stroke; AIDS; allergy; placenta; reproductive;
KW CGDD; cell growth; cell differentiation; cell death; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003027263-A2.
XX
PD 03-APR-2003.
XX
PF 26-SEP-2002; 2002WO-US031095.
XX
PR 28-SEP-2001; 2001US-0326389P.
PR 05-OCT-2001; 2001US-0327380P.
PR 05-OCT-2001; 2001US-0328188P.
PR 12-OCT-2001; 2001US-0329690P.
PR 26-OCT-2001; 2001US-0345384P.
PR 26-OCT-2001; 2001US-0348165P.
PR 02-NOV-2001; 2001US-0350219P.
PR 09-NOV-2001; 2001US-0344518P.
PR 09-NOV-2001; 2001US-0345143P.
PR 16-NOV-2001; 2001US-0332375P.
PR 03-DEC-2001; 2001US-0336908P.
PR 07-DEC-2001; 2001US-0340747P.
XX
PA (INCYTE GENOMICS INC.
XX
PI Azinza Y, Baughn MR, Becha SD, Borowsky ML, Chawla NK;
PI Elliott VS, Emerling BW, Gandhi AK, Gietzen KJ, Gorvad AE;
PI Griffin JA, Hafala AJA, Ison CH, Kable AE, Kalafus DP;
PI Lehr-Mason PM, Lu DAM, Marquis JP, Nguyen DB, Rankumar J;
PI Richardson TW, Sapperstein SK, Swarnakar A, Tang VT, Tran UK;
PI Warren BA, Xu Y, Yao MG, Yue H, Yue H;
XX
DR WPI. 2003-421159/39.
DR P-PSDB; ABR69612.
XX

PT New human proteins associated with cell growth, differentiation, and
PT death (CGDD), useful for diagnosing, treating and preventing diseases or
PT conditions associated with the aberrant CGDD expression e.g. cancer,
PT AIDS, or epilepsy.
XX
PS Claim 5; Page 317-318; 350pp; English.
XX
CC The invention relates to an isolated polypeptide associated with cell
CC growth, differentiation and death (CGDD). Also disclosed are the
CC polynucleotides encoding the polypeptides. The polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing
CC diseases or conditions associated with the decreased expression or over
CC expression of CGDD. Such diseases include cell proliferative (e.g.
CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
CC reproductive disorders, or disorders of the placenta. They are also
CC useful in assessing the effects of exogenous compounds on the expression
CC of nucleic acid and amino acid sequences of CGDD. The CGDD or its
CC fragments are useful in screening compounds for effectiveness as an
CC agonist or antagonist of the polypeptides, or in altering the expression
CC of the target polynucleotide and compounds that specifically bind to or
CC modulate the activity of the polypeptide. Microarrays consisting
CC polynucleotides of the invention are useful in monitoring or measuring
CC protein-protein interactions, drug-target interactions, and gene
CC expression profiles. Sequences given in records ACC90578-ACC90634
CC represent polynucleotides encoding CGDD proteins of the invention
XX
SQ Sequence 6962 BP; 2330 A; 1246 C; 1573 G; 1813 T; 0 U; 0 Other;
Query Match 16.1%; Score 32.4; DB 7; Length 6962;
Best Local Similarity 56.6%; Pred. No. 3.8; Mismatches 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 73 GCCTTTTAAAGGGGTGTCTCACTCTTCAGGGAAGTGGGAAAGGGGATCTGATTATT 132
DB 6835 GCTTTTAAAGGGGTTCAGATCCTGGAAAGAAATGCACAATAGTACTATGAGATG 6894
QY 133 GAGGTGTGAAGGAATAAATATATCATGTCACCAATAAACAACACTGT 178
DB 6895 TAGAGTGGCATCTGAAAACAATGTGTACACAAAGTAAACTGGCAGT 6940

RESULT 4
AAK84657
ID AAK84657 standard; DNA; 5983 BP.
AC AAK84657;
XX
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39469.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180528P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-02151135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230377P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231444P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233423P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250191P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
PS Disclosure; SEQ ID NO 39469; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to

CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 5983 BP; 1422 A; 1843 C; 1514 G; 1204 T; 0 U; 0 Other;

Query Match 15.7%; Score 31.6; DB 4; Length 5983;
Best Local Similarity 56.9%; Pred. No. 6.7;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 2 CAATGTTCTCCCTGAAGCCCTTCTCCCTGCCACACAGACGCACTTCCAAATCT 61
DB 772 CAAAATCCCTCCAGTCTACTCCCTCCATCCCATGCGCCACCTCTCCACACA 831
QY 62 GCCGCTGTTAGCTTGTAAAGGGGTCTCTACTCTCTTCAG 103
DB 832 GCCCTGCTTCTCTGCTGAATGAGCGTCTCCACCTCTG 873

RESULT 5
ABX81870
ID ABX81870 standard; cDNA; 296 BP.
AC ABX81870;
XX
DT 24-APR-2003 (first entry)
XX
DE Corn ear-derived polynucleotide (cpd) #330.
XX
KW Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;
KW structural gene; functional gene; regulatory gene;
KW corn ear-specific profile; gene transcription; gene expression;
KW hybrid plant; desirable trait expression; plant breeding program;
KW inheritance; desired characteristic; growth; development;
KW disease resistance; environmental adaptability; quality; yield;
KW multigene trait; plant; gene; ss.
XX
OS Zea mays.
XX
PN US6476212-B1.
XX
PD 05-NOV-2002.
XX
PF 14-MAY-1999; 99US-003113294.
XX
PR 26-MAY-1998; 98US-0086722P.
XX
PA (INCYTE) INCYTE GENOMICS INC.
XX
PI Lalgudi RV, Ito LY, Sherman BK;
XX
DR WPI; 2003-208840/20.
XX
PT Novel purified corn-ear derived polynucleotide useful as hybridization
PT probe for detecting polynucleotide in sample, and for identifying,
PT evaluating, and altering desired characteristics associated with growth,
PT development.
XX
PS Example; SEQ ID NO 330; 390pp; English.
XX

CC The present invention relates to the isolation of corn ear-derived
CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022
CC and SATMON023. Some of the cdps uniquely identify structural, functional,
CC and regulatory genes of corn ear. The polynucleotides sequences are
CC useful for detecting cdps in a sample, for producing a corn ear-specific
CC profile of gene transcription, for detecting altered gene expression in
CC inbred or hybrid plants, and for screening several molecules for specific

CC binding to the polynucleotide. The cdps are useful to identify, isolate,
CC or extend identical or related corn-ear nucleic acid sequences from DNA
CC libraries, and in nucleic acid amplification or hybridisation techniques
CC to follow the expression of desirable traits through plant breeding
CC programs. Preferably, the cdps are used to identify, evaluate, alter, or
CC follow the inheritance of desired characteristics associated with growth
CC and development, disease resistance, environmental adaptability, quality,
CC and yield of corn. The cdps are also useful as molecular markers for
CC studying inheritance and multigene traits in a plant breeding program.
CC The cdps are useful for producing purified corn-ear polypeptides by
CC recombinant techniques. They are also useful in diagnostic assays to
CC detect or confirm conditions or diseases associated with abnormal levels
CC of cdp expression. ABX81541-ABX89140 represent corn ear-derived
CC polynucleotides (cdps) of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site at
CC seqdata.uspto.gov/psipds/entry.html
XX

SQ Sequence 296 BP; 73 A; 59 C; 83 G; 78 T; 0 U; 3 Other;
Query Match 15.6%; Score 31.4; DB 7; Length 296;
Best Local Similarity 55.7%; Pred. No. 2.1;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 91 CTCACCTCTTCAGGAAAGTGGGAAAGGGATCTGATTATTGAGGTGCGAAGGATAA 150
DB 76 CTTACATACACCGTCCAGTCCGACAGGAGAGCAGTATCTGTTGCTGTCTATGC 135
QY 151 ATATCATGTCACACATATTAACAACTGTCGCGGATCTCTTAGAGGGA 196
DB 136 TTTATGAGCCCATATATACACAGTGTGAGNAGACTAGAAAGGA 181

RESULT 6
AAK70746
ID AAK70746 standard; DNA; 22700 BP.
AC AAK70746;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25558.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.

FT	variation	/standard_name= "Single nucleotide polymorphism" replace(25167,T) /tag= bb
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(25623,T) /tag= bc
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(27143,T) /tag= bd
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(27211,G) /tag= be
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(27393,A) /tag= bf
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(27553,T) /tag= bg
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(29141,G) /tag= bh
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(30549,T) /tag= bi
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(30984,T) /tag= bj
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(31266,A) /tag= bk
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(31425,C) /tag= bl
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(32198,G) /tag= bm
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(33129,C) /tag= bn
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(33292,G) /tag= bo
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(34124,T) /tag= bp
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(34668,A) /tag= bq
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(34807,T) /tag= br
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(36025,C) /tag= bs
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(36058,C) /tag= bt
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(36061,C) /tag= bu
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(36418,C) /tag= bv
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(36743,G) /tag= bw
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(36958,C) /tag= bx
FT	variation	/standard_name= "Single nucleotide polymorphism" replace(37577,G) /tag= by

Query Match	15.5%;	Score 31.2;	DB 7;	Length 213456;
Best Local Similarity	52.3%;	Pred. No. 45;		
Matches	69;	Conservative	0;	Mismatches 63; Indels 0; Gaps 0;
QY	8	TTCTCCCTCAAGCCCTCTTCCCTGCCCAACAGACGACGAACTTCCAAAATTCGCCCGT	67	
DB	64946	TTCTCCCTTACCAGTATTCGTAGAACCATATGCCAGTTAACTCCACCATTTTGCCTCT	65005	
QY	68	GTTTAGCCCTGTTAAAGGGGTCTCTCATCTCTTACGGGAAAGTGGGAAAAGGGGATCTGA	127	
DB	65006	TAGTGTGCATGCTTGAGTCCTCATTCACCCAGTTTCTTGAGATATTGGGAAAATGCGATCACC	65065	
QY	128	TTATTGAGCTGT	139	
DB	65066	AGTTTCAGGTTT	65077	
RESULT 8				
ABZ11734				
ID	ABZ11734	standard; cDNA; 1830 BP.		
XX	AC	ABZ11734;		
XX	DT	20-JAN-2003 (first entry)		
XX	DE	Human polynucleotide SEQ ID NO 616.		
XX	KW	Human; genome mapping; gene therapy; food supplement; virus; fungus;		
XX	KW	cell-proliferative disorder; neurodegenerative disease; bacterial;		
XX	KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;		
XX	KW	multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;		
XX	KW	arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;		
XX	KW	antiparkinsonian; antidiabetic; immunosuppressive; dermatological;		
XX	KW	haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;		
XX	KW	antiarthritic; Gene; ss.		
XX	OS	Homo sapiens.		
XX	OS	WO200270539-A2.		
XX	FN	12-SEP-2002.		
XX	PD	05-MAR-2002; 2002WO-US005095.		
XX	PF	05-MAR-2001; 2001US-00799451.		
XX	PR	(HYSE-) HYSEQ INC.		
XX	PA	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;		
XX	PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;		
XX	PI	Wehrman T, Wang J, Wang D, Drmanac RT;		
XX	XX	WPI; 2002-759812/82.		
DR	DR	P-PSDB; ABP69517.		
XX	XX	New polynucleotides comprising sequences assembled from expressed		
PT	PT	sequence tags (ESTs), useful for treating cell-proliferative,		
PT	PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet		
PT	PT	or coagulation disorders.		
XX	XX	Claim 1; SEQ ID NO 616; 1012pp + Sequence Listing; English.		
XX	PS	The invention relates to an isolated polynucleotide (I) comprising a		
XX	CC	nucleotide sequence selected from any of 948 sequences (ABZ1119-		
XX	CC	ABZ12066) or their mature protein coding portion, active domain coding		
XX	CC	protein or complementary sequences. The polynucleotides are useful for		
XX	CC	identifying expressed genes or for physical mapping of human genome. The		
XX	CC	encoded polypeptides (ABP69502-ABP69849) are useful as molecular weight		
XX	CC	markers, as a food supplement, for generating antibodies, in medical		
XX	CC	imaging, screening and diagnostic assays and for treating cell-		
XX	CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's		
XX	CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,		
XX	CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,		

CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
 CC arthritis, etc. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at fip.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1830 BP; 643 A; 283 C; 357 G; 547 T; 0 U; 0 Other;

Query Match 15.4%; Score 31; DB 6; Length 1830;
 Best Local Similarity 56.3%; Pred. No. 6.4;
 Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 76 TTGTTAAAGGGGTGCTCACTCTTCAGGGAAGTGGGAAAGGGGATCTGATTATTGAG 135
 DB 640 TTTTAAAGGGGGTTGAGTCTCTGGAAAGAAATGCACATAGTCACTATGAGATGAG 699
 QY 136 GTGTGAAGGAATAAATAATCATGTCACAAATAAACAACACTGT 178
 DB 700 AAGTGGCATCTGAACAATAATTGGTACACAAGTAAACTGGCAGT 742

RESULT 9

Continuation (4 of 5) of AAD53223 from base 300001 (Human chromosome 3 p-arm breakpoint
 WP Sequence split into 5 fragments LOCUS AAD53223 Accession AAD53223

WP	Fragment Name	Begin	End
WP	AAD53223_0	1	110000
WP	AAD53223_1	100001	210000
WP	AAD53223_2	200001	310000
WP	AAD53223_3	300001	410000
WP	AAD53223_4	400001	487980

Query Match 15.4%; Score 31; DB 7; Length 110000;
 Best Local Similarity 53.8%; Pred. No. 39;
 Matches 64; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 57 ATTCTGCCGTGTTAGCCCTTTAAAGGGGTCTCACTCTTCAGGAAAGTGGAAA 116
 DB 96672 AGTTGGCTAAATTGAATCTGTCAAGGGTTTCCCTTACTTGTGGCCAAATGATT 96731
 QY 117 AGGGGATCTGATTATTGAGGTGGGAAGGAATAAATAATCAGTCCACAAATAAACAAC 175
 DB 96732 TGTGGATTCTCCATGAGTTAGAGGCCAGGAATAATGAGCCAGGAAGATGAAC 96790

RESULT 10

ABX41793/C
 ID ABX41793 standard; cDNA; 360 BP.

AC ABX41793;
 XX
 XX
 DT 20-FEB-2003 (first entry)
 XX
 DE Bovine EST associated with lactation/muscle/fat deposition #6958.
 XX
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 XX muscle deposition; fat deposition; genome mapping; gene identification;
 KW Gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137139-A1.
 XX
 PD 26-SEP-2002.
 XX
 XX 24-SEP-2001; 2001US-00960352.
 XX
 PR 12-JAN-1999; 99US-0115707P.
 PR
 PR 11-JAN-2000; 2000US-00480902.
 PR
 PA (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.

(WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;
 XX WPI; 2003-110599/10.

PT New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and analysis,
 XX cattle breeding, or for genetically improving cattle.

PS Claim 2; SEQ ID NO 6958; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
 CC appearing as ABX34836-ABX49947, or complements of them. Also included are
 CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non-translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139

XX Sequence 360 BP; 118 A; 69 C; 47 G; 126 T; 0 U; 0 Other;

Query Match 15.3%; Score 30.8; DB 7; Length 360;
 Best Local Similarity 55.7%; Pred. No. 3.7;
 Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 54 AAAATTCGCCGTTTAGCTTGTAAAGGGGTCTCACTCTTCAGGAAAGTGGG 113
 DB 347 AAAATTTTATTCTCATTAAACTTTTTTAAATCGGTATAAAATTTTGACAGATTTTGG 288
 QY 114 AAAAGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAATCAAT 159
 DB 287 AAAAGTTCTTTTCATTAAAAAGTACTGATTTTAAAACTAATAATT 242

RESULT 11

ABX30387
 ID ABS30387 standard; DNA; 491 BP.

XX AC ABS30387;
 XX
 XX
 DT 25-FEB-2003 (first entry)
 XX
 DE Human liver single exon probe, SEQ ID No 5377.
 XX
 KW Human; single exon nucleic acid probe; liver; cirrhosis;
 KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
 KW coronary heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157273-A2.
 PN
 PD 09-AUG-2001.

```

XX PF 30-JAN-2001; 2001WO-US0000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488998/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX PS Claim 1; SEQ ID NO 5377; 658pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. AB25011-AB251005 represent human
XX CC liver single exon nucleic acid probes of the invention. Note: The
XX CC sequence information for this patent does not appear in the printed
XX CC specification but was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 491 BP; 135 A; 110 C; 144 G; 102 T; 0 U; 0 Other;

Query Match 15.1%; Score 30.4; DB 4; Length 491;
Best Local Similarity 71.4%; Pred. No. 5.8;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 119 GGGATCTGATTATTGAGGTGGAGGAAATAAATATCATCTCCACAAATAACAAA 174
DB 1 GGGATCTGATTATTGAGGATTTAGGAAGTAAGAAAGAACTTAAATAATATAGAAA 56

RESULT 12
AA56241
ID AAX56241 standard; DNA; 3574 BP.
XX AC AAX56241;
XX DT 16-JUL-1999 (first entry)
XX DE Mouse JMY nucleotide sequence.
XX KW Mouse; JMY; co-activator; p300/CBP; cell cycle; ss.
XX OS Mus sp.
XX PN WO9920752-A1.
XX XX
XX PD 29-APR-1999.
XX PF 21-OCT-1998; 98WO-GB003152.
XX PR 21-OCT-1997; 97GB-00022238.
XX PR 20-AUG-1998; 98GB-00018235.
XX XX
XX PA (UNIU ) UNIV GLASGOW.
XX CC

PI La Thangue NB, Shikama N;
XX DR WPI; 1999-302738/25.
XX DR P-PSDB; AAY09513.
XX PT JMY, a co-activator for p300/CBP.
XX PS Claim 8; Page 72-74; 81pp; English.
XX CC The present sequence encodes mouse JMY, which is a co-activator of
XX CC p300/CBP. The JMY polynucleotide (I), its complements or fragments (such
XX CC as primers) can be used to detect the presence of a JMY gene in a sample.
XX CC Modulators of JMY can be used to regulate the cell cycle. JMY antibodies
XX CC can be used to detect these modulators
XX SQ Sequence 3574 BP; 954 A; 905 C; 1029 G; 586 T; 0 U; 0 Other;

Query Match 15.1%; Score 30.4; DB 2; Length 3574;
Best Local Similarity 52.3%; Pred. No. 14;
Matches 67; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 9 TCTCCCTGAAGCCCTCTTCCCTGCCCAACACGACGACCACTTCCAAAATTTCTGCCCGTG 68
DB 2520 TCTTCTCCCAACACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2579
QY 69 TTATAGCTTTGTTAAAGGGGTGTCTCACTCTTCAAGGAAAGTGGGAAAGGGGATCTGAT 128
DB 2580 TGTTCAGAGGACAAATGGGGCGCTCCACCACTGCAGAGACACTGGAGAAAGATGCACCTTAG 2639
QY 129 TATTGAGG 136
DB 2640 GACGGAGG 2647

RESULT 13
AA72296/C
ID AAA72296 standard; DNA; 1466 BP.
XX AC AAA72296;
XX DT 06-DEC-2000 (first entry)
XX DE Drosophila odorant receptor DOR 47E.1 genomic DNA.
XX KW Odorant receptor; Drosophila; olfactory receptor;
XX KW G protein-coupled receptor; GPCR superfamily; transgenic insect;
XX KW insect behaviour modification; pest control; pollinator attraction;
XX KW biosensor; odour detection; odour identification; apiculture; ds.
XX OS Drosophila melanogaster.
XX PN WO200043410-A2.
XX PD 27-JUL-2000.
XX PF 25-JAN-2000; 2000WO-US001823.
XX PR 25-JAN-1999; 99US-0117132P.
XX PA (UYVA ) UNIV YALE.
XX PI Carlson JR, Kim J, Clyne FJ, Warr CG;
XX DR WPI; 2000-543246/49.
XX DR P-PSDB; AAB20912.
XX PT New nucleic acid encoding a Drosophila olfactory receptor, useful for
XX PT identifying modulating agents.
XX PS Example 1; Page 300-301; 303pp; English.
XX CC
XX CC Sequences AAA72285-A72298 represent genomic DNA encoding the Drosophila
XX CC melanogaster odorant receptors AAB20901-B20914. Drosophila odorant

```

CC receptor proteins (AAB20901-B20949) function as olfactory receptors, and
 CC are thought to be members of the G protein-coupled receptor (GPCR)
 CC superfamily, which is characterised by the presence of 7 transmembrane
 CC helices. Nucleic acids encoding the Drosophila odorant receptors may be
 CC used to generate expression constructs, host cells containing such
 CC constructs, and transgenic insects. Cells which express the odorant
 CC receptor genes may be used in methods to identify agents which modulate
 CC expression of these genes, and in methods to identify receptor binding
 CC partners. The Drosophila odorant receptor nucleic acids may also be used
 CC to identify corresponding genes in other insects, such as those which
 CC damage crops or transmit disease. The odorant receptor proteins may be
 CC used to identify agents which modulate their activity, to identify
 CC binding partners, as antigens to raise antibodies, and in methods to
 CC modify insect behaviour. The proteins may be also be used in methods of
 CC behaviour modification. Such methods may be used to study or modify
 CC of insect behaviour in response to odorants such as pheromones. Modification
 CC control (e.g., by disrupting the feeding or mating behaviours of pest
 CC species), or for enhancing plant pollination (by attracting pollinator
 CC to identify appetite suppressants, and/or nucleotides may also be used
 CC biosensors for the detection of explosives, drugs, perfumes or
 CC pollutants, and in apiculture to modify the behaviour of bees, for
 CC example, to increase the production of royal jelly
 CC
 SQ Sequence 1466 BP; 352 A; 341 C; 324 G; 449 T; 0 U; 0 Other;
 Query Match 15.0%; Score 30.2; DB 3; Length 1466;
 Best Local Similarity 55.1%; Pred. No. 11;
 Matches 59; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
 QY 48 ACTTCCAAATTCGCCGGTGTAGCTTTAGCTTTAAAGGGGTGTCTCACTCTTCAGGGAA 107
 DB 141 ACTTCAGAAACTGTCTCATGTGCAAGGGTTAAATCGCGCTTCACACTAAGTCAAGTGT 82
 QY 108 AGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAATAA 154
 DB 81 GTTCAGTTCAGGGGTTTAAACCTTTCAATTGAGTAGGAATAATGA 35
 RESULT 14
 AAA94812/c
 ID AAA94812 standard; DNA; 2075 BP.
 XX AAA94812;
 DT 23-FEB-2001 (first entry)
 XX Drosophila melanogaster odorant receptor DOR24 coding sequence.
 DE Odorant receptor; fruit fly; DOR24; odour recognition; pest control; ds.
 XX Drosophila melanogaster.
 XX Key Location/Qualifiers
 FH 1. .2075
 FT CDS /*tag= a
 FT /product= "DOR24"
 FT /note= "this sequence is interrupted by 3 introns"
 FT intron 187. .739
 FT /*tag= b
 FT /number= 1
 FT /cons_splice= (5'site:NO,3'site:NO)
 FT intron 1037. .1338
 FT /*tag= c
 FT /number= 2
 FT /cons_splice= (5'site:NO,3'site:NO)
 FT intron 1954. .2024
 FT /*tag= d
 FT /number= 3
 FT /cons_splice= (5'site:NO,3'site:YES)
 XX WO200050566-A2.

XX 31-AUG-2000.
 PD 25-FEB-2000; 2000WO-US004995.
 XX 25-FEB-1999; 99US-00257706.
 PR (UYCO) UNIV COLUMBIA NEW YORK.
 XX Voeshall LB, Amrein HO, Axel R;
 PI WPI; 2000-572081/53.
 XX P-PSDB; AAB26412.
 DR Novel nucleic acid encoding an insect odorant receptor, for identifying
 PT modulator compounds that are useful in controlling pest population.
 XX Claim 4; Fig 8; 176pp; English.
 XX The present sequence is the coding sequence for the Drosophila
 CC melanogaster odorant receptor DOR24. It was isolated using a differential
 CC cloning strategy. The odorant genes and proteins are useful as they aid
 CC in the study of the olfactory organ in mammals, as well as aiding the
 CC understanding of the link between odour recognition and behaviour in
 CC insects. They also enable the identification of compounds capable of
 CC activating and inhibiting the receptors, allow the control of pest
 CC populations via the use of alarm odour ligands and via the use of ligands
 CC which interfere with the interaction between odorant ligands and
 CC receptors associated with fertility
 XX Sequence 2075 BP; 548 A; 460 C; 447 G; 620 T; 0 U; 0 Other;
 SQ Query Match 15.0%; Score 30.2; DB 3; Length 2075;
 Best Local Similarity 55.1%; Pred. No. 13;
 Matches 59; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
 QY 48 ACTTCCAAATTCGCCGGTGTAGCTTTAGCTTTAAAGGGGTGTCTCACTCTTCAGGGAA 107
 DB 750 ACTTCAGAAACTGTCTCATGTGCAAGGGTTAAATCGCGCTTCACACTAAGTCAAGTGT 691
 QY 108 AGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAATAA 154
 DB 690 GTTCAGTTCAGGGGTTTAAACCTTTCAATTGAGTAGGAATAATGA 644
 RESULT 15
 ABL22180
 ID ABL22180 standard; DNA; 3548 BP.
 XX ABL22180;
 AC ABL22180;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 18013.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 09:55:14 ; Search time 39.5358 Seconds
(without alignments)
2821.370 Million cell updates/sec

Title: US-09-963-285-1_COPY_223_423

Perfect score: 201
Sequence: 1 acaaatgtctccctgaagc.....ggattccatagagggaaggag 201

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued Patents NA:*
- 1: /cgn2_6/prodata/2/ina/5A-COMB.seq:*
 - 2: /cgn2_6/prodata/2/ina/5B-COMB.seq:*
 - 3: /cgn2_6/prodata/2/ina/6A-COMB.seq:*
 - 4: /cgn2_6/prodata/2/ina/6B-COMB.seq:*
 - 5: /cgn2_6/prodata/2/ina/PCUS-COMB.seq:*
 - 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.4	15.6	296	4	US-09-313-294A-330
2	30.4	15.1	3574	4	US-09-311-236-1
3	30.2	15.0	1466	4	US-09-491-577-110
4	29.8	14.8	274	4	US-09-313-294A-4645
5	29.4	14.6	1191	4	US-09-252-991A-14425
6	29.4	14.6	640681	4	US-09-790-988-1
7	29.2	14.5	513	1	US-08-031-143B-70
8	29.2	14.5	1664976	4	US-08-916-421B-1
9	28.6	14.2	2920	4	US-08-976-259-10
10	28.2	14.0	3436	4	US-09-221-017B-335
11	28.2	14.0	1664976	4	US-08-916-421B-1
12	28	13.9	239	4	US-09-439-313-239
13	28	13.9	239	4	US-09-352-616A-239
14	28	13.9	239	4	US-09-232-149A-239
15	28	13.9	239	4	US-09-159-812-239
16	28	13.9	239	4	US-09-636-215-239
17	28	13.9	239	4	US-09-885-166A-239
18	28	13.9	239	4	US-09-688-489-239
19	27.8	13.8	5829	4	US-08-809-254A-4
20	27.8	13.8	41100	4	US-09-755-665-46
21	27.2	13.5	832	4	US-09-621-976-2813
22	27.2	13.5	2402	4	US-09-023-655-1317
23	27.2	13.5	2430	2	US-08-488-199-3
24	27.2	13.5	4797	4	US-09-643-597-134
25	27.2	13.5	4797	4	US-09-480-884A-134
26	27.2	13.5	4797	4	US-09-542-615A-134
27	27.2	13.5	4797	4	US-09-606-421B-134

Sequence 134, App
Sequence 3680, Ap
Sequence 1542, Ap
Sequence 5, Appli
Sequence 14, Appli
Sequence 1, Appli
Sequence 1682, Ap
Sequence 895, App
Sequence 28, Appl
Sequence 28, Appl
Sequence 3, Appli
Sequence 374, App
Sequence 4012, Ap
Sequence 2682, Ap
Sequence 389, App
Sequence 4413, Ap
Sequence 394, App
Sequence 11, Appli

28 27.2 13.5 4797 4 US-09-221-107-134
29 27 13.4 400 4 US-08-956-171E-3680
30 27 13.4 621 4 US-09-107-532A-1542
31 26.8 13.3 2628 4 US-09-294-531B-5
32 26.8 13.3 7218 1 US-08-232-463-14
33 26.8 13.3 35081 2 US-08-752-760A-1
34 26.6 13.2 468 4 US-09-543-681A-1682
35 26.6 13.2 1089 4 US-09-543-681A-895
36 26.6 13.2 2119 3 US-09-018-635-28
37 26.6 13.2 2119 4 US-09-912-962-28
38 26.6 13.2 2695 4 US-09-706-197-3
39 26.6 13.2 2712 4 US-09-976-594-374
40 26.4 13.1 400 4 US-08-956-171E-4012
41 26.4 13.1 1005 4 US-09-107-532A-2682
42 26.4 13.1 1249 4 US-08-956-171E-389
43 26.2 13.0 315 4 US-08-956-171E-4413
44 26.2 13.0 543 4 US-09-134-000C-394
45 26.2 13.0 1408 2 US-08-440-845D-11

ALIGNMENTS

RESULT 1
US-09-313-294A-330
; Sequence 330, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalngudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 330
; LENGTH: 296
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700549010H1
; NAME/KEY: unsure
; LOCATION: 22, 64, 167
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-330

Query Match 15.6%; Score 31.4; DB 4; Length 296;
Best Local Similarity 55.7%; Pred. No. 0.065;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 91 CTCACCTCTTCAGGGAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAGGAATAA 150
76 CTTACATACACCGTCCAGTCCGACAGGAGAGAGGTAATACCTGTTGTTCTGTCTGTC 135

QY 151 ATATCTGTCACAAATAAACAACAACTGCCGGGATCTCTAGAGGGA 196
DB 136 TTTATGAGCCCGAGATATAGCACAGTGTCTAGNAGACTAGAAGGGA 181

RESULT 2
US-09-311-236-1
; Sequence 1, Application US/09311236
; Patent No. 6376215
; GENERAL INFORMATION:
; APPLICANT: Lathangue, Nicholas B
; APPLICANT: Shikama, No. 6376215iko
; APPLICANT: The University Court of the University of Glasgow
; TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
; TITLE OF INVENTION: JMY, and uses thereof
; FILE REFERENCE: AHB/LF5731500

```

: CURRENT APPLICATION NUMBER: US/09/311,236
: CURRENT FILING DATE: 1999-05-13
: EARLIER APPLICATION NUMBER: GB 3722238.4
: EARLIER FILING DATE: 1997-10-21
: EARLIER APPLICATION NUMBER: GB 3918235.5
: EARLIER FILING DATE: 1998-08-20
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3574
: TYPE: DNA
: ORGANISM: Mus sp.
US-09-311-236-1

```

	Query Match	15.1%;	Score 30.4;	DB 4;	Length 3574;
	Best Local Similarity	52.3%;	Pred. No. 0.66;		
	Matches 67;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;
Qy	9	ACTCCCTGAAGCCCTCTTCCTGCGCCCAACGACACGACCAACTTCCAAAATTCGCGCGTG	68		
Db	2520	TCCTTCCTCCACACCTCCCTCTCCCTCCACCCACCCCTCCACCCCAACCCCTGCC	2579		
Qy	69	TTTAGCCCTTGTTAAAGGGGTGTCTCACTTCCTCAGGGAAAAGTGGAAAAGGGATCTGAT	128		
Db	2580	TGTTGCAAGGACAAATGGGGCTCCACCCTGCGACAGACACTGGAGAAAGATGCACTTAG	2639		
Qy	129	TATTGAGG	136		
Db	2640	GACGGAGG	2647		

RESULT 3

```

US-09-491-577-110/c
; Sequence 110, Application US/09491577
; Patent No. 6610511
; GENERAL INFORMATION:
; APPLICANT: Yale University
; APPLICANT: Carlson, John R.
; APPLICANT: Kim, Hunhyong
; APPLICANT: Clyne, Peter J.
; APPLICANT: Warr, Coral G.
; TITLE OF INVENTION: No. 6610511el Family of Odorant Receptor Genes in Drosophila
; FILE REFERENCE: 44574-5061-US
; CURRENT APPLICATION NUMBER: US/09/491,577
; CURRENT FILING DATE: 2000-01-25
; EARLIER APPLICATION NUMBER: US 60/117,132
; EARLIER FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 110
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: DOR 47E.1, NCBI Accession No. 6610511 AF156880
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1)..(121)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (423)..(478)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (669)..(728)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1345)..(1415)
; US-09-491-577-110

```

```
Query Match      15.0%; Score 30.2; DB 4; Length 1466;
Best Local Similarity 55.1%; Pred. No. 0.47;
Matches 59; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
```

Qy	48	ACTCTCAAAATTCGCGCGTGTATAGCTTGTGTTAAAGGGGTGTCCTACTCTTCAGGGAA	107
Db	141	ACTTCGAGAAACTGCTTCAATCGACAAGGGTTAATTCGGCCCTCACACTAAGTCAGTGT	82
Qy	108	ACTGGGAAAAGGGGATCTGATTTATGAGGTGTGGAAGGAATAATAA	154
Db	81	GTTTCAGTTCAGGGGTGTTTAAACCTTTCAATTGACTAGGAATAAATGA	35

RESULT 4

```

US-09-313-294A-4645
; Sequence 4645, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 4645
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700348780H1
; NAME/KEY: unsure
; LOCATION: 70, 129
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4645

```

Query Match 14.8%; Score 29.8; DB 4; Length 274;
Best Local Similarity 54.2%; Pred. No. 0.25;
Matches 58; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

91 CTCACCTCTCCAGGAAAGTCGGAAAGGGGATCTGATTATTTCAGCTGTGGAGGGAATAA 150
65 CTTACNTACACGGTCCAGGTCCGACAGGAGAGCAGGTAAATCTGGTGTTGCTGTCTATGC 124
151 ATAATCAGTCCACAAATAAAACAACTGTCCGGGATTCCTAGAGGGAA 197
125 TTTTATGACCCCGATATAGTATACAGTGTTCACGACCTAGAAAGGAA 171

RESULT 5

```

US-09-252-991A-14425
, Sequence 14425: Application US/09252991A
, Patent No. 6551795
, GENERAL INFORMATION:
, APPLICANT: Marc J. Rubenfield et al.
, TITLE OF INVENTION: NUCLEIC ACID AND AMINO
, TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSIS
, FILE REFERENCE: 107196.136
, CURRENT APPLICATION NUMBER: US/09/252,991A
, PRIORITY FILING DATE: 1999-02-18
, PRIOR APPLICATION NUMBER: US 60/074,788
, PRIOR FILING DATE: 1998-02-18
, PRIOR APPLICATION NUMBER: US 60/094,190
, PRIOR FILING DATE: 1998-07-27
, NUMBER OF SEQ ID NOS: 33142
, SEQ ID NO 14425
, LENGTH: 1191
, TYPE: DNA
, ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14425

```

Query Match	14.6%	Score 29.4;	DB 4;	Length 1191;
Best Local Similarity	58.6%;	Pred. No. 0.82;		
Matches	51;	Conservative	0;	Mismatches 36;
			Indels	0;
			Caps	0;

RESULT 7
US-08-031-143B-70/c
; Sequence 70, Application US/08031143B
; Patent No. 5518860
; GENERAL INFORMATION:
; APPLICANT: LEONARD, WARREN J.; NOGUCHI, MASAYUKI;
; APPLICANT: MCBRIDE, O. WESLEY
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF XSCID
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORD PERFECT # 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/031,143B
; FILING DATE: 12-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

RESULT 8
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequen
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)

```

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664855)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Query Match 14.5%; Score 29.2; DB 4; Length 1664976;

Best Local Similarity 57.8%; Pred. No. 60;

Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY

76 TTGTTAAAGGGGTGTCTCACTCCTTCAGGAAAGGGGATCTGATTATTGAG 135

Db 793847 TTGATTAAGTATGTTGTTATCTCAGGTAAGAGTTAAATCCCTATATCAATAAAACAG 793906
Qy 136 GTCTGAAGGAATAAATAATCAGTCACAA 165
Db 793907 GTTTGTCAACAATATGTTATAGTCCAAAA 793936
RESULT 9
US-08-976-259-10
; Sequence 10, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,698
; REFERENCE/DOCKET NUMBER: 1489.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-976-259-10
Query Match 14.2%; Score 28.6; DB 4; Length 2920;
Best Local Similarity 55.6%; Pred. No. 2.8;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Qy 99 TTCAGGAAGTGGAAAGGGATGCTGATTATGAGGTGCGAAGGAATAAATATCAG 158
Db 1902 TTCAGTATATGAGTAATGAATGAATATACACAGACATTAAGAAAAATATCAG 1961
Qy 159 TCACAAATAAACAACCTGCCGGGATCTCAGAGGAA 197
Db 1962 ACACCAAAAAACAAGACCCGCATATCTCTTCGGGAA 2000
RESULT 10
US-09-221-017B-335
; Sequence 335, Application US/09221017B
; Patent No. 644799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 335:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3436 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...3436
; US-09-221-017B-335
Query Match 14.0%; Score 28.2; DB 4; Length 3436;
Best Local Similarity 50.4%; Pred. No. 4.3;
Matches 69; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
Qy 50 TTCCAAAATTCGCCCGTGTTCAGCTTTGTTAAAGGGGTGTCCTCCTTCAGGAAAG 109
Db 2667 TCGAAGTTCGTTCGACTTTTAGGTCCGGAATTTCTCGTTTACAGTCAGCAAA 2726
Qy 110 TGGGAAAAGGGGATCTGATTATTCAGGTGTGGAAGGAATAAATAATCAGTCCCAATATA 169
Db 2727 CGTGTTCGAGAAAATTTTGTGTTGGCGTGGAAAAAATAATTTATCGAACCAACGAA 2786
Qy 170 ACAACTGTCGGGATT 186
Db 2787 AAAAATACGCCGAGT 2803
RESULT 11
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729

GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/09/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319236)..(319236)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g

```
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
;
Query Match 14.0%; Score 28.2; DB 4; Length 1664976;
Best Local Similarity 55.7%; Pred. No. 1.2e+02;
Matches 54; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 14 CTGAGCCCTCTCCCTGCCACACGAGCACTTCCAAAATTCGCCGCTGTAG 73
DB 1193490 CTGATAATGCTTCTCCACTCCCGTAGTGAGCAATTAACATCTCTCCCTTCATGAA 1193431

QY 74 CCTTGTAAGGGGTCTCTCACTCTTCAGGAAAGT 110
DB 1193430 TCTCTTAATTCATATCTTTGGTGGAGTAAAGT 1193394

RESULT 12
US-09-439-313-239
; Sequence 239, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-239

Query Match 13.9%; Score 28; DB 4; Length 239;
Best Local Similarity 48.7%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 41 ACCAGCAACTTCCAAAATTCGCCGCTGTAGCTTTAAAGGGGTGTCTCACTCCTT 100
DB 38 AACATAAAGTTACATAACTGCTTCTGTCAAACCATGATCTGAGCTTTGTGACACCC 97
QY 101 CAGGGAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTC 160
DB 98 AGAATAAATTAAGAGAGGCAACATAATACCTTAGATCAAGAAACATTACACAGTT 157
QY 161 CACAAATAAACAAACTGTCGCCGATTCCTAGAGGGA 196
DB 158 CAACGTGTTAAAAATAGCTCAACATTGAGCCAGTGA 193

RESULT 13
US-09-352-616A-239
; Sequence 239, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqi
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-352-616A-239

Query Match 13.9%; Score 28; DB 4; Length 239;
Best Local Similarity 48.7%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 41 ACCAGCAACTTCCAAAATTCGCCGCTGTAGCTTTAAAGGGGTGTCTCACTCCTT 100
DB 38 AACATAAAGTTACATAACTGCTTCTGTCAAACCATGATCTGAGCTTTGTGACACCC 97
QY 101 CAGGGAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTC 160
DB 98 AGAATAAATTAAGAGAGGCAACATAATACCTTAGATCAAGAAACATTACACAGTT 157
QY 161 CACAAATAAACAAACTGTCGCCGATTCCTAGAGGGA 196
DB 158 CAACGTGTTAAAAATAGCTCAACATTGAGCCAGTGA 193

RESULT 14
US-09-232-149A-239
; Sequence 239, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-232-149A-239

Query Match 13.9%; Score 28; DB 4; Length 239;
Best Local Similarity 48.7%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

; ORGANISM: Homo sapien
US-09-232-149A-239
Query Match 13.9%; Score 28; DB 4; Length 239;
Best Local Similarity 48.7%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 41 ACCAGCAACTTCCAAAATTCTGCCGCTGTTAGCCTTTAAAGGGTGTCTCACTCCTT 100
Db 38 ACATAAAAGTTCACTAATCTGCTCTCAACCATGATGACTGCTTTGTGACACCC 97
QY 101 CAGGAAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTC 160
Db 98 AGAATAAATAAGAGAGGCAACATAATACCTTAGAGATCAAGAAACATTACACAGTT 157
QY 161 CACAATAAACAACACTTCCGGGATTCCTAGAGGA 196
Db 158 CAACGTGTTAAAAATAGCTCAACATTCAGCCAGTGA 193

RESULT 15
US-09-159-812-239
; Sequence 239; Application US/09159812A
; Patent No. 6613872
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF
; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.428C5
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-159-812-239

Query Match 13.9%; Score 28; DB 4; Length 239;
Best Local Similarity 48.7%; Pred. No. 1.1;
Matches 76; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 41 ACCAGCAACTTCCAAAATTCTGCCGCTGTTAGCCTTTAAAGGGTGTCTCACTCCTT 100
Db 38 AACATAAAAGTTCACTAATCTGCTCTCAACCATGATGACTGCTTTGTGACACCC 97
QY 101 CAGGAAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTC 160
Db 98 AGAATAAATAAGAGAGGCAACATAATACCTTAGAGATCAAGAAACATTACACAGTT 157
QY 161 CACAATAAACAACACTGTCGGGATTCCTAGAGGA 196
Db 158 CAACGTGTTAAAAATAGCTCAACATTCAGCCAGTGA 193

Search completed: March 25, 2004, 15:34:33
Job time : 43.5358 secs

QY 51 TCCCGTGTAGCTTGTAAAGGGGTCTCACTCTTCAAGGAAAGTGGGAAAGG 120
Db 283 TCCCGGTGTAGCTTGTAAAGGGGTCTCACTCTTCAAGGAAAGTGGGAAAGG 342
QY 121 GATCTGATTATTAGCTGTGAAGGAATAAATAATCAGTCCACAAATAAACAACACTGTCC 180
Db 343 GATCTGATTATTAGCTGTGAAGGAATAAATAATCAGTCCACAAATAAACAACACTGTCC 402
QY 181 GGGATTCCTAGGGAAGGAG 201
Db 403 GGGATTCCTAGGGAAGGAG 423

RESULT 2

US-09-963-285-5
; Sequence 5, Application US/09963285
; Patent No. US20020090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerback, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6021
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1649)...(438)
US-09-963-285-5

Query Match 32.8%; Score 66; DB 9; Length 6021;

Best Local Similarity 73.1%; Pred. No. 2.2e-11;
Matches 155; Conservative 0; Mismatches 45; Indels 12; Gaps 5;

QY 1 ACAATGTTCTCCCTGAAGCCCTCTTCCCTGCCCAACAGACGACGCAACTTCCAAAATTC 60
Db 195 ACAATGTTTATCTGTGCGCTTCTTCCCTACCCACCGGACCAACTTCCAGAGGT 254
QY 61 TCCCGGTG---TTTAGCCTTGTAAAGGGGTCTCACTCTTCAAGGAAAGTGGGAAA 117
Db 255 TCTGCGAGCATAGAGCCATTCGTTAGGACATCTCGGTGCTTCTGAGGAAGCGGACCGA 314
QY 118 G---GGATCTGATTA---TTGAGTGTGGAAGGATAAATAATCAGTCCACAAATAAAC 171
Db 315 GCAGGATCCGATGACGACTGGAGATGTTGAGGATAAAT-ACCAGTCCACAAATAAAC 373
QY 172 AAACGTG---CCGGATTCCTAGGGAAGGAG 201
Db 374 AAACGTGCCCCGGATTCCTAGGGAAGGAG 405

RESULT 3

US-10-424-599-39980/c
; Sequence 39980, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 39980
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_136101C.1
US-10-424-599-39980

Query Match 17.7%; Score 35.6; DB 12; Length 477;

Best Local Similarity 51.2%; Pred. No. 0.13; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 79;

QY 36 ACCAGACCAGCAACTTCCAAAATTCGCCGCTGTAGCTTTGTTAAAGGGGTGTCTCAC 95
Db 383 ACAACAACAATGAATAACAACAAGGAATGATTTTCATCTTCGAAAATTAATGACTTAA 324
QY 96 TCCTTCAGGGAAGTGGGAAGGGGATCTGATTTAGGTGTGGAAGGAATAATAAT 155
Db 323 ACATTCATGAATCAGGATCGAGGATCTTATTATAAATTAATTGACACAATTAATAAT 264
QY 156 CAGTCCACAAATAAACAACAACCTGTCGGGATTCCTAGAGGGAA 197
Db 263 TGAARAAACCAAAATTTGTACAGTACGGTATCCAAAGACTGAA 222

RESULT 4

US-10-229-834A-6/c
; Sequence 6, Application US/10229834A
; Publication No. US2003015003A1
; GENERAL INFORMATION:
; APPLICANT: Lawrence Berkeley National Laboratory
; APPLICANT: Rubin, Edward
; APPLICANT: Pennacchio, Len
; TITLE OF INVENTION: A novel apolipoprotein gene involved in lipid metabolism
; FILE REFERENCE: IB-1709
; CURRENT APPLICATION NUMBER: US/10/229,834A
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 60/318,219
; PRIOR FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 263744
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (750)..(850)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1584)..(1683)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2154)..(2154)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2358)..(2457)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3208)..(3307)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc feature

LOCATION: (4044)..(4143)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (4809)..(4908)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (4921)..(4921)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (5652)..(5751)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6469)..(6469)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6477)..(6477)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6482)..(6482)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6487)..(6487)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6489)..(6490)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6523)..(6622)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (6637)..(6637)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (7410)..(7509)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (8177)..(8177)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (8248)..(8347)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (8358)..(8358)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (8967)..(8967)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (9057)..(9057)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (9105)..(9204)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (9980)..(10079)

OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (10092)..(10092)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (10827)..(10926)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (11701)..(11800)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (11816)..(11816)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (11827)..(11827)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (11829)..(11829)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (12565)..(12664)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (13327)..(13327)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (13440)..(13539)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (13554)..(13554)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (14133)..(14133)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (14213)..(14213)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (14293)..(14392)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (14559)..(14559)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (14900)..(14900)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (15119)..(15218)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (15231)..(15233)
OTHER INFORMATION: gap of unknown length
FEATURE:
NAME/KEY: misc feature
LOCATION: (15919)..(15919)
OTHER INFORMATION: gap of unknown length


```

; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53)13/B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32307
; LENGTH: 2443
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73291G02_FLI
US-10-425-114-32307

Query Match      15.3%; Score 30.8; DB 12; Length 2443;
Best Local Similarity 55.7%; Pred. No. 11;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 91 CTCACCTCTTCAGGAAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAA 150
DB 1425 CTTACATACCGGTGAGTCCGACAGGAGAGCAGGTAATACCTGGTGTCTGTCATGC 1484

QY 151 ATAATCAGTCCACAATAAACAACACTGTCGGGATTCCTAGAGGA 196
DB 1485 TTTATGAGCCCAAGATATAGCACAGTGTGACGAGACTAGAAAGGGA 1530

RESULT 9
US-09-864-761-14243
; Sequence 14243, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14243
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009893.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92
US-09-864-761-14243

Query Match      15.1%; Score 30.4; DB 9; Length 491;
Best Local Similarity 71.4%; Pred. No. 7.5;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 119 GGGATCTGATTATTGAGGTGTGGAAGGAATAATAATCATCTCCACAATAAACAAA 174
DB 1 GGGATCTGATTATTGAGGTGTGGAAGGAATAATAATCTTAAATAATATAGAAA 56

RESULT 10
US-09-912-917-1
; Sequence 1, Application US/09912917
; Patent No. US20020028487A1
; GENERAL INFORMATION:
; APPLICANT: Lathangue, Nicholas B
; APPLICANT: Shikama, No US2002028487A1iko
; TITLE OF INVENTION: The University Court of the University of Glasgow
; TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
; FILE REFERENCE: JMY and uses thereof
; FILE REFERENCE: AHB/LP5731500
; CURRENT APPLICATION NUMBER: US/09/912,917
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 09/311,236
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: GB 9818235.5
; PRIOR FILING DATE: 1998-08-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3574
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-912-917-1

Query Match      15.1%; Score 30.4; DB 9; Length 3574;
Best Local Similarity 52.3%; Pred. No. 18;
Matches 67; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 9 TCTCCCTGAAGCCCTCTTCTCCCTGCCCAACCCAGACAGCACTTCCAAAATTCGCCCGTG 68
DB 2520 TCTTCTCCCAACACCTCCCTCCCTCCCAACCCCAACCCCAACCCCAACCCCTGCC 2579

QY 69 TTTAGCCCTTGTAAAGGGGTGTCTCACTCTTCAGGAAAGTGGGAAAGGGATCTGAT 128
DB 2580 TGTTCGAAGGACATGCGGCTCCCACTGCAGACACACTGAGAGAGATGACACTTAG 2639

QY 129 TATTGAGG 136
DB 2640 GACGGAGG 2647

```

```
RESULT 11
US-09-911-077A-14/c
; Sequence 14, Application US/09911077A
; Publication No. US20030114399A1
; GENERAL INFORMATION:
; APPLICANT: BLAKELY, RANDY D.
; APPLICANT: APERSONDARAM, SUBRAMANIAM
; APPLICANT: FERGUSON, SHAWN
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA
; FILE REFERENCE: VBLT:008US
; CURRENT APPLICATION NUMBER: US/09/911,077A
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 14
; LENGTH: 142299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1305)..(127835)
; OTHER INFORMATION: N = A, C, G or T/U
US-09-911-077A-14

Query Match      14.7%; Score 29.6; DB 10; Length 142299;
Best Local Similarity 57.6%; Pred. No. 1.7e+02;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 11 TCCCTGAAGCCCTCTTCCTGCCAACCCAGACCACTTCCAAATTTCTGCCCGTGTT 70
Db 42105 TCTCGGTGCTACACCACTTCTTAATGAACCTTGACCACTTCTGTCATTC 42046

QY 71 TAGCCTTTGTAAGGGGTGTCATCTCTTCA 102
Db 42045 CATCCTTTGTTGAGTGGTGATAGTGTTTA 42014

RESULT 12
US-10-027-632-194870
; Sequence 194870, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194870
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-194870
```

```
Query Match      14.6%; Score 29.4; DB 15; Length 624;
Best Local Similarity 54.1%; Pred. No. 18;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 39 AGACCAGCAACTTCCAAATTTCTGCCCGTGTATTACCTTTGTAAAGGGGTGTCACCTCC 98
Db 28 AGAAGAGTATTTCACACCTTCTTCTGTGTTTCCAGGGTAAGGGGTGTCCTCATTTA 87

QY 99 TTCAGGGAAGTGGGAAAGGGGATCTGATTTATTTGAGGTGTGGAAGGAATA 149
Db 88 CTGAGTACCACACACAGGGGGTGAATCTGAATGGGACTTGGGAATGTAGA 138

RESULT 13
US-10-027-632-98041
; Sequence 98041, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98041
; LENGTH: 1680
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-98041

Query Match      14.6%; Score 29.4; DB 15; Length 1680;
Best Local Similarity 60.8%; Pred. No. 28;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 114 AAAAGGGGATCTGATTTATTTGAGGTGGGAAGGAATAATAATCAGTCCCAATAAACAA 173
Db 1169 AAAAGAAGATCTGTTCTTATGAGAGGCTAGAGTACAGCAGTGCAGTGAAAAATAT 1228

QY 174 ACTGTCGGGATTCCTAGA 192
Db 1229 TTAATCCTGAATACTTAGA 1247

RESULT 14
US-10-027-632-98042
; Sequence 98042, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

QY 174 ACTGTCGGGATTCCTAGA 192
DB 1229 TTAATCCTGAATACTAGA 1247

Search completed: March 25, 2004, 15:52:59
Job time : 178.116 secs

PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 98043
LENGTH: 1680
TYPE: DNA
ORGANISM: Human
US-10-027-632-98042

Query Match 14.6%; Score 29.4; DB 15; Length 1680;
Best Local Similarity 60.8%; Pred.No. 28;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 114 AAAAGGGATCTGATTATGAGTGTGGAGGAATAATATCACTCCACAAATAACAA 173
DB 1169 AAAAGAGATCTGTTCTTATGAGAGGCTAGAGTACAGCAGTCAGTCAAAATAT 1228

QY 174 ACTGTCGGGATTCCTAGA 192
DB 1229 TTAATCCTGAATACTAGA 1247

RESULT 15
US-10-027-632-98043
Sequence 98043, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 98043
LENGTH: 1680
TYPE: DNA
ORGANISM: Human
US-10-027-632-98043

Query Match 14.6%; Score 29.4; DB 15; Length 1680;
Best Local Similarity 60.8%; Pred.No. 28;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 114 AAAAGGGATCTGATTATGAGTGTGGAGGAATAATATCACTCCACAAATAACAA 173
DB 1169 AAAAGAGATCTGTTCTTATGAGAGGCTAGAGTACAGCAGTCAGTCAAAATAT 1228

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 09:53:09 ; Search time 1518.97 Seconds
(without alignments)
3951.570 Million cell updates/sec

Title: US-09-963-285-1_COPY_223_423
Perfect score: 201
Sequence: 1 acaaatgtctccctgaagc.....ggattccctagagggaagag 201

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB	ID Description
C 1	79	39.3	644	28 B2249455
C 2	43	21.4	362	28 BH858390
C 3	36.8	18.3	1002	13 BU538230
C 4	35.6	17.7	755	14 CD438041

5	35.6	17.7	1068	9	AL527496
6	35.4	17.5	699	29	CE672514
7	35.2	17.5	875	14	CD327143
8	35.2	17.5	982	23	CNS0142R
9	35.2	17.5	1004	29	CNS006TX
10	34.2	17.0	1201	9	AL525275
11	34	16.9	607	14	CB981228
12	34	16.9	717	10	BE391709
13	34	16.9	995	28	CC292246
14	33.8	16.8	604	14	CB237820
15	33.4	16.6	702	28	AZ847437
16	33.2	16.5	313	29	AG263957
17	33.2	16.5	489	28	AZ160863
18	33.2	16.5	714	13	EX509001
19	33.2	16.5	730	12	BJ014817
20	33.2	16.5	787	13	BU537610
21	33	16.4	539	28	AZ092244
22	32.8	16.3	331	9	AI414501
23	32.8	16.3	355	14	CD969718
24	32.8	16.3	479	13	BY240517
25	32.8	16.3	616	28	AZ396298
26	32.8	16.3	710	28	CC310470
27	32.6	16.2	399	14	R81779
28	32.6	16.2	480	28	AZ329846
29	32.6	16.2	563	14	AZ515154
30	32.6	16.2	371	9	AA587556
31	32.4	16.1	518	28	AQ792699
32	32.4	16.1	576	13	EX673778
33	32.4	16.1	703	28	CC166880
34	32.4	16.1	834	14	CB948166
35	32.4	16.1	909	28	CC346613
36	32.4	16.1	1201	9	AL548542
37	32.2	16.0	304	10	BB221567
38	32.2	16.0	344	29	CE639367
39	32.2	16.0	390	14	CB778000
40	32.2	16.0	435	29	CE543124
41	32.2	16.0	462	14	CB606450
42	32.2	16.0	470	12	EG375690
43	32.2	16.0	538	14	CB717464
44	32.2	16.0	538	14	CB717464
45	32.2	16.0	563	14	CB609216

ALIGNMENTS

RESULT 1
BZ249455/C

LOCUS CH230-314P20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-314P20, genomic survey sequence.
ACCESSION BZ249455
VERSION BZ249455.1
KEYWORDS GI:23910687
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 644)

AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-314P20.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	79	39.3	644	28	B2249455
C 2	43	21.4	362	28	BH858390
C 3	36.8	18.3	1002	13	BU538230
C 4	35.6	17.7	755	14	CD438041

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 314 row: P column: 20
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers

FEATURES

1. .644
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SnHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-314P20"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pPARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SnHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 39.3%; Score 79; DB 28; Length 644;
Best Local Similarity 73.9%; Pred. No. 1.le-11;
Matches 156; Conservative 0; Mismatches 45; Indels 10; Gaps 4;
QY 1 ACAATGTTCTCCCTGAAGCCCTCTCCCTCCCAACACGACGACCTCCAAATTC 60
DB 325 ACACATGTTTATCTGTGCGCTTCTCCCTCCCAACGCGACCAACGACTTCCAGAGGT 266
QY 61 TGCCCGTG---TTTAGCCTTGTAAAGGGGTCTCACTCTCCAGGAAAGTGGGAAA 117
DB 265 TCTACAGGATAGAGCCCTTCCTAGGAGCTCTCAGTCTCTGAGGAGCGGACGAG 206
QY 118 G-GGATCTGATTA---TTGAGGTGTGAAGGAATAAATAATCAGTCCCAATAAACA 172
DB 205 CAGGATCCGATGAGACTGCGAGGTGTGAAGGAATAAATAACGAGTCCCAATAAACA 146
QY 173 AACTGTCC---GGATTCCTAGAGGAGGAG 201
DB 145 AACTGTCCGGGATCTCTAGAGGAGGAG 115

RESULT 2

BH858390/c
LOCUS BH858390.1 362 bp DNA linear GSS 13-NOV-2002
DEFINITION B5_216a2_t7 Mouse Retroviral Tagged Cancer Gene Database Mus
musculus genomic clone B5_216a2, genomic survey sequence.
ACCESSION BH858390
VERSION BH858390.1 GI:21709211
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 362)
AUTHORS Suzuki,T., Shen,H., Akagi,K., Morse,H.C., Malley,J.D., Naiman,D.Q.,
Jenkins,N.A. and Copeland,N.G.
TITLE New genes involved in cancer identified by retroviral tagging
JOURNAL Nat. Genet. 32 (1), 166-174 (2002)
MEDLINE 22194816
PUBMED 12185365
COMMENT Contact: Copeland NG
Mouse Cancer Genetics Program
National Cancer Institute
Bldg. 539, Rm. 229, Frederick, MD 21702-1201, USA
Tel: 301 846 1260
Fax: 301 846 6666
Email: copeland@ncifcrf.gov
Class: PCR with specific primers.
Location/Qualifiers

FEATURES

source

1. .362
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="B5_216a2"
/sex="female"
/tissue type="leukemia"
/clone_lib="Mouse Retroviral Tagged Cancer Gene Database"
/note="Inverse PCR method
(http://genome2.ncifcrf.gov/RtCGD)"

ORIGIN

Query Match 21.4%; Score 43; DB 28; Length 362;
Best Local Similarity 67.0%; Pred. No. 0.15;
Matches 77; Conservative 0; Mismatches 35; Indels 3; Gaps 1;
QY 1 ACAATGTTCTCCCTGAAGCCCTCTCCCTCCCAACACGACGACCTCCAAATTC 60
DB 124 ACAATGTTTATCTGTGCGCTTCTCCCTCCCAACGCGACCACTTCCAGAGGT 65
QY 61 TGCCCGTG---TTTAGCCTTGTAAAGGGGTCTCACTCTCCAGGAAAGTGG 112
DB 64 TCTCGAGGATAGAGCCATCCGTAGGAGACATCTCGGTCTCTGAGGAAGCG 10

RESULT 3

BUS38230
LOCUS BUS38230 1002 bp mRNA linear EST 13-SEP-2002
DEFINITION AGENCOURT 10186385 NIH MGC 107 Homo sapiens cDNA clone
IMAGE:6568756 5', mRNA sequence.
ACCESSION BUS38230
VERSION BUS38230.1 GI:22848671
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1002)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2754 row: p column: 04
High quality sequence stop: 102.
Location/Qualifiers

FEATURES

1. .1002
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6568756"
/tissue type="adenocarcinoma, cell line"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 18.3%; Score 36.8; DB 13; Length 1002;
Best Local Similarity 63.6%; Pred. No. 9;

```

Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 8 TTTCCTCCGTAAGCCCTCTTCCTCCCAACAGACGACACTTCCAAAATTTGCGCCGT 67
    |||||
Db 795 TTTTCCCGGAAGCCCGTTTACGGGGCCCGGGACCCCACTTACACGGTTTCCCGGT 854
    |||||
QY 68 GTTTAGCCCTCTTTAAAGGGGTGTTTCAC 95
    |||||
Db 855 TTTTCAATTTTACAAGGGGTTTAAAC 882
    |||||

RESULT 4
CD438041 755 bp mRNA linear EST 03-JUN-2003
LOCUS EL01N0508607.b EndospERM_5 Zea mays cDNA, mRNA sequence.
DEFINITION CD438041
ACCESSION CD438041
VERSION CD438041.1 GI:31353684
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 755)
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE Sequencing of the maize endospERM ESTs
JOURNAL Unpublished (2002)
COMMENT Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.

FEATURES
source
1..755
location/Qualifiers
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/cultivar="W22"
/tissue_type="EndospERM of 7-23DAP"
/clone_lib="EndospERM_5"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 17.7%; Score 35.6; DB 14; Length 755;
Best Local Similarity 58.5%; Pred. No. 19;
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 91 CTCACCTCTTCAGGGAAGTGGGAAAGGGGATCTGATTATGAGTGTGGAAGGAATAA 150
Db 225 CTTACATACACGGTCAGTTCGGACGAGGAGCAGGTAATACTGGTGTGTCGTCGTC 284
QY 151 ATAATCAGTCCACAATAAACAACATGTCCTCCGGATCTCTAGAGGA 196
Db 285 TTTATGAGCCCAATATAGCAGGTGTGAGCATTTTAAAGGGA 330

RESULT 5
AL527496 1068 bp mRNA linear EST 23-MAY-2003
LOCUS AL527496 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DC024YG07 5-PRIME, mRNA sequence.
ACCESSION AL527496
VERSION AL527496.2 GI:31065347
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 1068)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12790989.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6496.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC024AD040P1&cluster=6496.f. Contact:
Peng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CS0DC024AD040P1.
Location/Qualifiers
1..1068
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC024YG07"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 17.7%; Score 35.6; DB 9; Length 1068;
Best Local Similarity 43.4%; Pred. No. 20;
Matches 33; Conservative 24; Mismatches 19; Indels 0; Gaps 0;
QY 103 GCGAAGTGGGAAAGGGGATCTGATTATGAGTGTGGAAGGAATAATCATCTCCA 162
Db 866 GGAAAGGGGAAAAAAATTTKWWTTKKKKKKKKKKKKKKKKKKKKKKKKKKKK 925
QY 163 CAAATAAACAAACTGT 178
Db 926 AAAAAAAAMAAAWKK 941

RESULT 6
CE672514/c 699 bp DNA linear GSS 29-SEP-2003
LOCUS tigr-gss-dog-17000329437427 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE672514
VERSION CE672514.1 GI:36991514
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 699)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Deicher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers

FEATURES

```



```
source
1. .699
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BatX1; Libraries were prepared from peripheral blood"

ORIGIN
Query Match 17.6%; Score 35.4; DB 29; Length 699;
Best Local Similarity 59.4%; Pred. No. 22;
Matches 60; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 64 CCGTGTTCCTTTGTTAAAGGGGTGCTCACTCTTCAGGAAAGTGGGAAAGGGGAT 123
Db 180 CTGCTTTGGTATACAAAGGGTGTGCTCTCGAGGAAACATAGGAAAGGGAA 121
QY 124 CTGATTATGAGGTGTGGAAGGAATAATATCATCAGTCCACA 164
Db 120 ATGGAAAGGTGGTATGAGGAAACAAATCCTCATTTACCA 80

RESULT 7
CD327143/c
LOCUS
DEFINITION
IMAGE:6947793 5', mRNA sequence.
ACCESSION
CD327143
VERSION
CD327143.1 GI:31091474
KEYWORDS
EST.
SOURCE
Xenopus laevis (African clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 875)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue procurement:
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLA14561 row: a column: 08
High quality sequence start: 11
High quality sequence stop: 683.
FEATURES
Location/Qualifiers
1. .875
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6947793"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC Eye1"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection
(XGC) library."

ORIGIN
Query Match 17.5%; Score 35.2; DB 14; Length 875;
Best Local Similarity 51.2%; Pred. No. 25;

source
1. .982
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN11A16"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : SP6"

ORIGIN
Query Match 17.5%; Score 35.2; DB 29; Length 982;
Best Local Similarity 33.5%; Pred. No. 26;
Matches 52; Conservative 39; Mismatches 64; Indels 0; Gaps 0;

QY 20 CCCTCTTCCTGCCCAACACAGACCAACTCCAAATTCGCCCGTTTAGCCTTGT 79
Db 828 CCCCCCCCCCCCCCCCCCGTGTGATTKKTKMGMDCTSKHGGWGTNCRKTATA 887
QY 80 TAAAGGGGTGTCCTCACTCTTCAGGAAAGTGGGAAAGGGGATCTGATTATTAGGTGT 139
Db 888 TAGGGDWTGAAMRRKRTGGGGGSRADGKKTGKKKGKGTGTGWTATATKKKKKK 947
QY 140 GGAAGGAATAATAATCAGTCCACAATAAACA 174
Db 948 KKKTKWAAAADAAYAAAKAAWAKATAWAAWA 982

RESULT 9
CNS006TX/c
```

LOCUS CNS006TX 1004 bp DNA linear GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR14N23 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL065920
 VERSION AL065920.1 GI:4944888
 KEYWORDS GSS
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1004)
 AUTHORS
 TITLE Direct Submission
 JOURNAL
 COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammosser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES
 Location/Qualifiers
 1..1004
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR14N23"
 /clone_lib="RPCI-98"
 /note="end : TET3"
 ORIGIN
 source
 Query Match 17.5%; Score 35.2; DB 29; Length 1004;
 Best Local Similarity 55.6%; Pred. No. 26;
 Matches 40; Conservative 11; Mismatches 21; Indels 0; Gaps 0;
 QY 103 GGGAAAGTGGGAAGGGGATCTGATTATTGAGTGTGGAGGATTAATAATCAGTCCA 162
 Db 483 GAGAGGGGGAAAKAAAKAAWKAATKTKATKATKGTGAAAGTAAANGAAGAAAA 424
 QY 163 CAAATAAACA 174
 Db 423 AAAAAA 412
 RESULT 10
 AL525275 1201 bp mRNA linear EST 22-MAY-2003
 LOCUS AL525275 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
 DEFINITION CDNA clone CSODC011YM07 3-PRIME, mRNA sequence.
 ACCESSION AL525275
 VERSION AL525275.2 GI:31043530
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1201)
 AUTHORS Li W.B., Gruber C., Jessee J., and Polayes D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12788768.

Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 4913.f For
 more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODC011AG04NP1&cluster=4913.f>. Contact :
 Feng Liang Email : liang@lifetech.com URL :
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>
 Paraday Avenue Genoscope sequence ID : CSODC011AG04NP1.
 FEATURES
 Location/Qualifiers
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSODC011YM07"
 /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Query Match 17.0%; Score 34.2; DB 9; Length 1201;
 Best Local Similarity 26.7%; Pred. No. 49;
 Matches 35; Conservative 49; Mismatches 47; Indels 0; Gaps 0;
 QY 15 TGAAGCCCTTCCCGCCACACAGACGACACTTCCAAATTCGCCCGTGTAGC 74
 Db 1032 TKTMMCCCCCCCM 1091
 QY 75 CTGTGTTAAAGGGGTGCTCACTCCCTTCAGGAAAGTGGGAAAGGGGATCTGATTATCA 134
 Db 1092 KKGKKGKKGKGGGGKKTCTTTATKGGGSAAGAAAGAAAGKTKATTKGKTGKG 1151
 QY 135 GGTGTGGAAGG 145
 Db 1152 GGGGGGRARDK 1162
 RESULT 11
 CB581228/c
 LOCUS CB581228 Rattus norvegicus
 DEFINITION nrhy7-00006-b1-A nrhy7 (10850) Rattus norvegicus CDNA clone
 CB581228
 ACCESSION CB581228
 VERSION CB581228.1 GI:29525269
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 607)
 AUTHORS Angen Est Program.
 TITLE Angen Rat EST Program
 JOURNAL Unpublished (2003)
 COMMENT Contact: Dan Fitzpatrick
 Angen, Inc
 One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
 Tel: 805 447-4881
 Plate: 00006 row: b column: 1.
 FEATURES
 Location/Qualifiers
 1..607
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="nrhy7-00006-b1"
 /clone_lib="nrhy7 (10850)"
 /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; W Rat

[illegible]

```

DEFINITION AGENCOURT 11477053 NIH_MGC_166 Mus musculus cDNA clone
IMAGE:30252220 5', mRNA sequence.
ACCESSION CB237820
VERSION CB237820.1 GI:28289398
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 604)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: csapbs-remail.nih.gov
Tissue Procurement: T. Rajendra Kumar and Martin M. Matzuk
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM18 row: j column: 05
High quality sequence stop: 194.
FEATURES
    source
    1..604
        /organism="Mus musculus"
        /mol_type="mRNA"
        /db_xref="taxon:10090"
        /clone="IMAGE:30252220"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: brain; Vector: pDNR-LIB; Site 1: SfiI
        (ggccattagggc); Site 2: SfiI (ggccgccctgggc);
        Non-normalized full-length enriched library 5' and 3'
        adaptors were used in cloning as follows: 5' adaptor
        sequence: 5'-CACGCGCATATGGC-3' and 3' adaptor sequence:
        5'-ATTCTAGGCGGGGGCGGACATG-dt(30)BN-3' (where B = A,
        C, or G and N = A, C, G, or T). Average insert size 2.05
        kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
        by PCR. This library was enriched for full-length clones
        and was constructed by Clontech Laboratories (Palo Alto,
        CA). Corp. "
```

ORIGIN

```

Query Match 16.8%; Score 33.8; DB 14; Length 604;
Best Local Similarity 52.5%; Pred. No. 62;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 42 CCAGCAACTCCAAATCTGCCCGCTTTAGCCTGTTTAAAGGGGTCTCACTCTTC 101
Db 388 CCCACACCTTTATAAAAAAACCCCTTTTATTCTTTTAAAGGGGGGGAACAGGG 447

Qy 102 AGGGAAGTGGGAAAGGGATCTGATTATTAGGTGTGGAAGGAATAATATCAGTCC 161
Db 448 GGGGTTAAAAAAGAGTGTTTTAAAGNAGGGGGGGGGGTATAAACCCCTTTT 507

Qy 162 ACAATTAACAACACTGTCGG 182
Db 508 AAAAAAATAATAGTGGGGG 528
```

RESULT 15

```

AZ847437 702 bp DNA linear GSS 21-FEB-2001
LOCUS 2M0148G04F Mouse 10kb plasmid UUCG1M library Mus musculus genomic
DEFINITION clone UUC2M0148G04 F, genomic survey sequence.
ACCESSION AZ847437
VERSION AZ847437.1 GI:13028699
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 702)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederkausen,A. and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308 Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0148 row: G column: 04
Seq primer: CGTGTAAACGACGCGCCAGT
Class: Plasmid ends
High quality sequence stop: 702.
FEATURES
    source
    1..702
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="UUC2M0148G04"
        /sex="Male"
        /lab_host="E. Coli strain XL10-Gold, TI-resistant, P-"
```

```

        /clone_lib="Mouse 10kb plasmid UUCG1M library"
        /note="Vector: PWD42mv; Purified genomic DNA from M.
        musculus C57BL/6J (male) was obtained from the Jackson
        Laboratory Mouse DNA Resource
        (http://www.jax.org/resources/documents/dnares/). The DNA
        was hydrodynamically sheared by repeated passage through a
        0.005 inch orifice at constant velocity. The sheared DNA
        was blunt end-repaired with T4 DNA polymerase and T4
        polynucleotide kinase. Adaptor oligonucleotides were
        ligated to the blunt ends in high molar excess. The
        adaptor DNA was purified and size-selected for a 9.5 to
        10.5 kb range using preparative agarose gel
        electrophoresis. Vector DNA was prepared from a derivative
        of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
        inducible derivative of plasmid R1. The vector was ligated
        with adaptors complementary to the insert adaptors and
        purified. The sheared, adaptor mouse DNA was annealed to
        adaptor vector DNA, and transformed into
        chemically-competent E. coli XL10-Gold (Stratagene) cells
        and selected for ampicillin resistance."
```

ORIGIN

```

Query Match 16.6%; Score 33.4; DB 28; Length 702;
Best Local Similarity 58.6%; Pred. No. 81;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 103 GGGAAAGTGGGAAAAGGGATCTGATTATTGAGGTGTGGAAGGAATAATATCAGTCCA 162
Db 193 GGGGAATGGGTAGACAGCAAGTAATATTCTGATGTAATGAATAATAATGAAAA 252

Qy 163 CAATATAACAACCTGTCGGGATTCCTAGAGGAGGAG 201
Db 253 AAAAAAGAAAAGTTTATGAACAACAAGAAAATGGG 291
```

```

Search completed: March 25, 2004, 15:30:23
Job time : 1523.97 secs
```